

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2002, 02:39:19 ; Search time 3112 Seconds  
(without alignments)  
7191.535 Million cell updates/sec

Title: US-09-555-342B-1\_COPY\_1733\_2501  
Perfect score: 769  
Sequence: 1 tcaactcgtgttcagagc.....attgaggagagcgaagca 769

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl :

- 1: gb\_ba.\*
- 2: gb\_hgt.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vt.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vt.\*
- 30: em\_hgt\_hum.\*
- 31: em\_hgt\_inv.\*
- 32: em\_hgt\_other.\*
- 33: em\_hgt\_mus.\*
- 34: em\_hgt\_pln.\*
- 35: em\_hgt\_rod.\*
- 36: em\_hgt\_mam.\*
- 37: em\_hgt\_vrt.\*
- 38: em\_sy.\*
- 39: em\_hgt\_hum.\*
- 40: em\_hgt\_mus.\*
- 41: em\_hgt\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	769	100.0	3442	9	AB008430	AB008430 Homo sapi
2	631.4	82.1	3902	10	BC030329	BC030329 Mus muscu
3	478.8	62.3	2632	10	BC004009	BC004009 Mus muscu
4	312.8	40.7	2431	10	BC027077	BC027077 Mus muscu
5	280.8	36.5	3997	9	AB018336	AB018336 Homo sapi
6	271.2	35.3	3719	10	BC009153	BC009153 Mus muscu
7	241	31.3	106578	9	AL137249	AL137249 Human DNA
8	204.2	26.6	224310	2	AC094777	AC094777 Rattus no
9	199.4	25.9	221502	2	AC122885	AC122885 Mus muscu
10	162.4	21.1	393	11	G60059	G60059 SHGC-130885
11	136.8	17.8	152053	2	AC022669	AC022669 Homo sapi
12	120.8	15.7	426	6	AX261839	AX261839 Sequence
13	119.2	15.5	190588	2	AC101725	AC101725 Mus muscu
14	119.2	15.5	221502	2	AC122885	AC122885 Mus muscu
15	111.8	14.5	149417	2	AC118779	AC118779 Rattus no
16	97.2	12.6	42014	2	AC018327	AC018327 Drosophila
17	97.2	12.6	160710	3	AC009537	AC009537 Drosophila
18	97.2	12.6	164713	3	AC007589	AC007589 Drosophila
19	97.2	12.6	303209	3	AE003604	AE003604 Drosophila
20	97	12.6	62537	2	AC101100	AC101100 Mus muscu
21	95.8	12.5	105605	2	AC115336	AC115336 Rattus no
22	95.8	12.5	176630	2	AC109427	AC109427 Rattus no
23	95	12.4	3055	10	BC026860	BC026860 Mus muscu
24	87.8	11.4	134911	9	AC005104	AC005104 Homo sapi
25	84.2	10.9	152053	2	AC022669	AC022669 Homo sapi
26	83.6	10.9	1286	6	AX411309	AX411309 Sequence
27	83.6	10.9	1821	9	AK002045	AK002045 Homo sapi
28	83.6	10.9	3610	6	AX405929	AX405929 Sequence
29	83.6	10.9	3842	9	AB037783	AB037783 Homo sapi
30	82	10.7	1478	9	BC013319	BC013319 Homo sapi
31	81.4	10.6	3220	5	AF017370	AF017370 Danio rer
32	80.4	10.5	120911	2	AC130919	AC130919 Rattus no
33	80.2	10.4	67896	2	AC131316	AC131316 Mus muscu
34	79.6	10.4	133745	2	AC131454	AC131454 Strongylo
35	73.8	9.6	2605	9	BC021301	BC021301 Homo sapi
36	72.8	9.5	462	6	AX371287	AX371287 Sequence
37	70.8	9.2	138025	9	HSDJ20N4	AL096709 Human DNA
38	65.8	8.6	2956	10	AF017369	AF017369 Mus muscu
39	64.4	8.4	1896	9	BC032232	BC032232 Homo sapi
40	64.4	8.4	3224	9	BC032429	BC032429 Homo sapi
41	64.4	8.4	4670	9	AK000004	AK000004 Homo sapi
42	62.2	8.1	2170	9	AK097217	AK097217 Homo sapi
43	62.2	8.1	120911	2	AC130919	AC130919 Rattus no
44	62.2	8.1	149417	2	AC118779	AC118779 Rattus no
45	62.2	8.1	224310	2	AC094777	AC094777 Rattus no

ALIGNMENTS

RESULT 1  
AB008430  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

AB008430  
Homo sapiens mRNA for CDEP, complete cds.  
AB008430  
AB008430.1 GI:2766164  
CDEP.  
Homo sapiens embryo cartilage chondrocyte cDNA to mRNA.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (sites)  
Koyano,Y., Kawamoto,T., Shen,M., Yan,W., Noshiro,M., Fujii,K. and  
Kato,Y.  
Molecular cloning and characterization of CDEP, a novel human

protein containing the ezrin-like domain of the band 4.1 superfamily and the Dbl homology domain of Rho guanine nucleotide exchange factors

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

98086358  
2 (bases 1 to 3442)  
Koyano, Y., Kawamoto, T. and Kato, Y.  
Direct Submission  
Submitted (22-OCT-1997) Takeshi Kawamoto, Hiroshima University  
School of Dentistry, Department of Biochemistry; 1-2-3 Kasumi  
Minami-ku, Hiroshima, Hiroshima 734, Japan  
(E-mail: tkawamo@ipc.hiroshima-u.ac.jp, Tel: 082-257-5688,  
Fax: 082-257-5629)

## FEATURES

source

1..3442  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="chondrocyte"  
/tissue\_type="cartilage"  
/dev\_stage="embryo"  
49..3186  
/function="Rho Guanine Nucleotide Exchange Factor"  
/note="Band 4.1 superfamily"

CDS

/codon\_start=1  
/product="CDEP"  
/protein\_id="BAA24267.1"  
/db\_xref="GI:2766165"  
/translation="MGTELEORPTGSRGAPENSIGSTLERGOKPPTPSKLYSIKI  
QMLDDTQAFEPVQAPGKVLDDVCHNLNVEGDIYFGLFDPHKKITVNDLKLPIV  
KQIRPRHVVVKYVVKFPDPDHTLOEELRYLFAIQVKDLAGRITCNDLSAALI  
SHVQSEIGDEALDEHLAKNKYIPQDALEDKIVFHHNHIGQTPAESDFOLLEI  
ARILEMYGIRLHPADRGRTKINLAVANTGLVFQGTKINAFNWKVRLSKFRKRF  
LILKRPDANSAYQDTLEPLMASRDFCKSWKICVHEHAFRLFEPEPKPKPVLFSRG  
SSFEFGTQKQVLDYKVEGCHKVQPERKHSKIHSLASQTELSNVELEQSOQS  
TSLFGEAESFGSGSRGKEPKVSAGEPSHESPPRRSPAGNKQDGAASPTSE  
EEVVKDRTQSKPQPPQSTGSLTSPHLSLSVNSQGVAPANVTLSPNLSPTDK  
ASPLSLNDAQCPRTDEDEGRKRPPTDKYFAKEVSTERTYKLDLEVTISWF  
QSTVSKEDMAPDFLSLIFPNFEPLEHFKPHTNFKLEIQRALWEGRSNAQRDYQNTFL  
DVLNKGIMGHLAAHLWKHEALEALENGILKSSRLLENFCRDFELQKVCYLPNTFL  
LRPLRLHMYQVLELCKKHPPSHADPRDCRAALAEITENVQALHGTMIKMFQKL  
HELKDLGINDLVVPGREFIRLGLSKLGGKGLQORFFLNFVLLTISRLGLTASND  
FKVHGQLGNTLVIESEDEWGVPHCLTLRGQROSIIVAASSRSEMEKRWEDIQMID  
LAEKSSPAPEFLASSPDNKSPEDEATAQDESEDSLSRSLERAPHRGNTMVH  
CWHRTSVSDFSTAVENQLSGLNLLRFRKNSGWQKLWVFTNFCLFFYKSHQDNHP  
LASLPLGLSYLTIPSESENIQKDYVFKLHFKSHVYVYFRAESEYTFERMEVIRSATSS  
ASRPHVLSHKSLEYV"

polyA\_site

BASE COUNT 864 a 952 c 927 g 699 t  
ORIGIN

Query Match 100.0%; Score 769; DB 9; Length 3442;  
Best Local Similarity 100.0%; Pred. No. 1e-175;  
Matches 769; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCTCGTGGTTTCAGACACAGTGGAGCAAGAGGAGCCATCGCGGAGCACTGAAA 60  
|||||  
Db 1733 TCACCTCGTGGTTTCAGACACAGTGGAGCAAGAGGAGCCATCGCGGAGCACTGAAA 1792  
QY 61 GTCTCATATTCGGAATTTGAACTTTCGACAAATTTCACTACTAATTTTCACAGGAAA 120  
|||||  
Db 1793 GTCTCATATTCGGAATTTGAACTTTCGACAAATTTCACTACTAATTTTCACAGGAAA 1852  
QY 121 TTGACACAGCACTTGCCCTGTGGAGGCGCTCAATGCCCAATTCAGAGATTACCCAAA 180  
|||||  
Db 1853 TTGACACAGCACTTGCCCTGTGGAGGCGCTCAATGCCCAATTCAGAGATTACCCAAA 1912  
QY 181 GAATCGGCGATGTCATGCTGAAGAACATTCAGGCGATGAAGCACCTGGCGGCTCACCTGT 240  
|||||  
Db 1913 GAATCGGCGATGTCATGCTGAAGAACATTCAGGCGATGAAGCACCTGGCGGCTCACCTGT 1972  
QY 241 GGAGCACAGCAGGCGCTTGGAGGCGCTGGAGATGGAATCAAGAGCTCCCGCGGCGTGG 300  
|||||

Db 1973 GGAAGCACAGCAGGCGCTTGGAGGCGCTTGGAGATGGAATCAAGAGCTCCCGCGGCTGG 2032  
QY 301 AGAAGCTTCGAGAGACTTTTGGAGTGCAGAAAGGTGTGTACCTACCGTCAACACCTTCC 360  
|||||  
Db 2033 AGAAGCTTCGAGAGACTTTTGGAGTGCAGAAAGGTGTGTACCTACCGTCAACACCTTCC 2092  
QY 361 TCCTCGGCGCACTGCACCGGCTCATGCACTCAAGCAGGTCTCTGGAGGCGGTGTGCAAAAC 420  
|||||  
Db 2093 TCCTCGGCGCACTGCACCGGCTCATGCACTCAAGCAGGTCTCTGGAGGCGGTGTGCAAAAC 2152  
QY 421 ACCACCCCGGAGGCGACCGGCTCAGGAGTCCGAGAGCGCTTGGCAGAGATCAGG 480  
|||||  
Db 2153 ACCACCCCGGAGGCGACCGGCTCAGGAGTCCGAGAGCGCTTGGCAGAGATCAGG 2212  
QY 481 AGATGGTGGCAGCTCCACCGTACGATGATCAAGATGAGAAATTTCCAGAGCTGCACG 540  
|||||  
Db 2213 AGATGGTGGCAGCTCCACCGTACGATGATCAAGATGAGAAATTTCCAGAGCTGCACG 2272  
QY 541 AACTCAAGAAAGATTTGATTGGCATTGACAATCTTGTGTTCCGGGAAGGAGTTTCATCC 600  
|||||  
Db 2273 AACTCAAGAAAGATTTGATTGGCATTGACAATCTTGTGTTCCGGGAAGGAGTTTCATCC 2332  
QY 601 GTCTGGGAGCGCTCAGCAAGCTCTCGGGAGAGGGCTCCAGAGCGCATGTTCTTCTCTGT 660  
|||||  
Db 2333 GTCTGGGAGCGCTCAGCAAGCTCTCGGGAGAGGGCTCCAGAGCGCATGTTCTTCTCTGT 2392  
QY 661 TCAACGAGCTCTGCTATACAGGCGGCGCTCACGCGCTCCCAATCAGTTTAAAGTCC 720  
|||||  
Db 2393 TCAACGAGCTCTGCTATACAGGCGGCGCTCACGCGCTCCCAATCAGTTTAAAGTCC 2452  
QY 721 ACGGCGAGCTCCGCTCTATGCGCATGACGATTGAGGAGGAGCAAGACGA 769  
|||||  
Db 2453 ACGGCGAGCTCCGCTCTATGCGCATGACGATTGAGGAGGAGCAAGACGA 2501  
RESULT 2  
BC030329 3902 bp mRNA linear ROD 07-AUG-2002  
LOCUS Mus musculus, clone IMAGE:5376197, mRNA, partial cds.  
DEFINITION BC030329  
ACCESSION BC030329  
VERSION BC030329.1 GI:20987935  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Scurionathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 3902)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL  
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov  
COMMENT Contact: MGC help desk  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: http://www.nisc.nih.gov/  
Contact: nisc\_mgc@nih.gov  
Akter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,  
Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,  
Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, E.D.





QY 630 AAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAAGCAGCTCCTGCTATACACAGCCGG 689  
 Db 241 AAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAAGCAGCTCCTGCTATACACAGCCGG 300  
 QY 690 GGGCTGACCGCTCCCAATCAGTTTAAAGTCAGGGCAGCTCCGCTCATGGCATGAGC 749  
 Db 301 GGACTGACCGCATCTAATCAGTTTAAAGTCAGGGCAGCTCCCACTCATGGCATGAGC 360  
 QY 750 ATTGAGGAGGAGGAGGAGCA 769  
 Db 361 ATCGAGGAGGAGGAGGAGCA 380

RESULT 5  
 LOCUS AB018336 3997 bp mRNA linear PRI 16-JUN-1999  
 DEFINITION Homo sapiens mRNA for KIAA0793 protein, complete cds.  
 ACCESSION AB018336  
 VERSION AB018336.1 GI:3882306  
 KEYWORDS  
 SOURCE Homo sapiens adult male brain cDNA to mRNA, clone\_lib:pBluescriptII  
 SK plus clone:hk05692.

ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 AUTHORS Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Miyajima,N.,  
 Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.  
 Prediction of the coding sequences of unidentified human genes. XI.  
 The complete sequences of 100 new cDNA clones from brain which code  
 for large proteins in vitro  
 DNA Res. 5 (5), 277-286 (1998)  
 99087487

TITLE  
 2 (bases 1 to 3997)  
 Ohara,O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.  
 Direct Submission

JOURNAL  
 MEDLINE Submitted (08-OCT-1998) Osamu Ohara, Kazusa DNA Research Institute,  
 Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba  
 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,  
 Fax:+81-438-52-3914)

FEATURES  
 Location/Qualifiers  
 1..3997  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="hk05692"  
 /sex="male"  
 /tissue\_type="brain"  
 /clone\_lib="pBluescriptII SK plus"  
 /dev\_stage="adult"  
 1..3997  
 /gene="KIAA0793"  
 118..3282  
 /gene="KIAA0793"  
 /codon\_start=1  
 /product="KIAA0793 protein"  
 /protein\_id="BAA34513.1"  
 /db\_xref="GI:3882307"  
 /translation="MGEIEGTVLQAGRLGATPVGVSTLEPGQTLPRMQEHL  
 HLRYLLDNTWEIDIEPKDQGVLLTOVKRLNLEVDYFGMEFQNTQYTWLEPM  
 RPIHQIRRRANVLRVAKFFPPDPGQLEQYTHRYLQALQRLDLLERLTQADTTA  
 ALLSHLQISGIDYDETLDRHLKNVEYLFQOQHLKLEKILEFHQKHVGTQPAESDFQ  
 VLETKARLEMGIRFHMVADREGTKIQLAVSHMGLVFQGTTKINTFNWVKRLKSF  
 RKRFILKLPHEVHPYQDTLEFLGSRDECKNFKICVEYHTFFLLDQPKPAKAVF  
 FSRGSSFRYSGRTOKQVDYFKDSGMKRIPIYERHSKTHTSVRALTADLPKQSLFPE  
 GLRTPASPSSANAYLSLSPSLTPSGLPFEKDDSSSLDTPQVSYVKSPPAERSGAVA  
 GGPDPAPQALGPALQPGSLSTKSPSSPSPLSLSPAFOVYKLPAPAEQSSPL  
 LSPVLDAAGGAGMOCDEPRHVRPADEAYFIVKEILLATERTYLLKEDVYVWFSAVV  
 KEDAMPATLMLFSNIDPIYEFHGRFLREVEQRLALWEGPSKAHTKGSHQRIIGILL  
 RNMQLKEFTSYFQRHDEVLTELEKATKCKLEAVYKEFELOKVCYLPPLATLKLPI  
 ORLLHYRLLLRGHYSFGPHDYADCHDAKLAITEVTTTQHLIRLENLQKLTELQ  
 RDLVGIENLIAPGRFIREGCLHLTKGLQORMFFLFSMDLLYTSKGVAGTSHFRIR  
 GLLPLQGLMVSPEENSWSPHCFTIYAAQKTIYVAASRTLEKEMWLDLSAIAQAKS  
 GGDTPALPGRVTCTRPSPRSPNEVSLQESEDARGVRSLSLEGHQHQRANTTMHVCYV

RNTSVSRADHSAAVENQLSGYLLRRKFNHGWQKLVVFTNFCFLFFYKTHDDYPLAS  
 LPLLGYSVSIIPREADGIHKDYVKLFQKSHVYFFRAESKYTFERWMEVIQGAASSAGR  
 APSIVQDGPQPPSSGLEGMVRGKEE"  
 BASE COUNT 998 a 1132 c 1068 g 799 t  
 ORIGIN

Query Match 36.5%; Score 280.8; DB 9; Length 3997;  
 Best Local Similarity 61.1%; Pred. No. 2.1e-57;  
 Matches 472; Conservative 0; Mismatches 297; Indels 3; Gaps 1;

QY 1 TCACCTTCGTGGTTTCAGAGCAGCTGAGCAAGAGAGCAGCCATGCGGGAAGCACTGAAGA 60  
 Db 1787 TTACCGTGTGGTTCCGACGCGCAGTGTGAAAGGAGGAGCCATGCTGCGACTCATGA 1846  
 QY 61 GTCCTCATATTCGGAATTTTGAACCTTTGCACAAATTTTCATACTAATTTTCTCAAGAAA 120  
 Db 1847 CGTGTGCTTCTTCCAACTCGATCCCATCTATGAGTTCCACAGAGGCTTCCTCGCGGAGG 1906  
 QY 121 TTGAGCAAGCAGCTTGCCCTGTGGAAAGGCGCTCAATGCCAAATCAGAG---ATTACC 177  
 Db 1907 TGGAGCAGAGGCTGGCAGCTCTGGAAAGGCGCTCCAAAGCCACACAAAGGCGAGTCATC 1966  
 QY 178 AAGAATCGCGCATGTCTGCTGAAGAACAATTCAGGGCATGAAGCACTGGCGGCTCACC 237  
 Db 1967 AAGCAATCGGGACATCTCTGCTCAGGAACATGCGCCAGTTAAAGGAGTTTACCAGCTACT 2026  
 QY 238 TGTGGAAGCACAGCAGGAGCGCTTGGAGCGCTTGAGAGTGAATCAAGACTCCCGCGGC 297  
 Db 2027 TCAAGACATGACGAGGCTTGAAGAACTTGAAAGGCTTACCAACGCTGTGAAGAAGT 2086  
 QY 298 TGGAGAACTTCTGACAGAGCTTTGAGCTGCAGAAAGTGTGTTACCTACCGCTCAACACCT 357  
 Db 2087 TGGAGCAGCTGTACAAAGGAGTTTGAGCTGCAGAAAGTGTGCTACTTGCCTCTCAACACGT 2146  
 QY 358 TCTCTCTGGGCGACTGACCGGCTCATGCACTACAGCAGTCTCTGGAGCGGCTGTGCA 417  
 Db 2147 TCTGTCTGAAGCCCATCCAGCGGCTGCTGCCTACCGCTGTGCTGCGCGGCTATGG 2206  
 QY 418 AACACCAACCGCGGAGCGGCGGCTTGCAGGAGCTGCGGAGCGCTTTGGGAGAGATCA 477  
 Db 2207 GACATTTACAGCCCGGCGGCGGCGGCTGAGTACGCTGACATGAGCCCTGAAAGCCATCA 2266  
 QY 478 CGGAGATGGTGGCACACTCCACGGTACGATGATGAGAGATTTCCAGAACTGC 537  
 Db 2267 CAGAGGTGACCAACACACTACAGCAGTCTCATCTCCGGCTGGAGAACCTCTCAGAGCTAA 2326  
 QY 538 AGAACTCAAGAAAGATTTGATTTGGCATTTGACAATCTTGTGGTTCCGGGAGGAGTTCA 597  
 Db 2327 CGGAGCTGACGCGGCGGCGGCTGTTGGGCTAGAGAACCTCATTTGCTCTGCGCAGGAGTTCA 2386  
 QY 598 TCCGCTGCGGCGGCGGCGGCTCTCGGGAAGGCGCTCCAGCAGCGCATCTTCTTCC 657  
 Db 2387 TCGTGAGGCGTGCCTTTCACAGCTCACCAGAAAGGCGCTGCGAGCAGAGATGTTTTTC 2446  
 QY 658 TGTTCACACGCTCTCTGCTATACACAGCGGCGGCTGACGGGCTCCAATCAGTTTAAAG 717  
 Db 2447 TGTTCACATATGTTGCTGTACACAAAGCAAGGAGTTTCAGGAGCACGCCACTTCCGGA 2506  
 QY 718 TCACGGGAGCTCCCGCTCTCTATGCGATGACGATTCAGGAGGAGCGGAGACGA 769  
 Db 2507 TCCGGGCGCTCTCTTCCCTTCCAGGCGATGCTGTGTGAAGAAAGTATAACGA 2558

RESULT 6  
 LOCUS BC009153 3719 bp mRNA linear ROD 07-AUG-2002  
 DEFINITION Mus musculus, similar to KIAA0793 gene product, clone MGC:6304  
 IMAGE:2655209, mRNA, complete cds.  
 ACCESSION BC009153  
 VERSION BC009153.1 GI:14318718  
 KEYWORDS MGC.  
 SOURCE house mouse.  
 ORGANISM Mus musculus



only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sv: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at  
<http://www.sanger.ac.uk/projects/C.elegans/wormpep>  
 This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr13>  
 RP11-111L24 is from the library RPC1-11.1 constructed by the group of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBACE3.6  
 IMPORTANT: This sequence is not the entire insert of clone RP11-111L24. It may be shorter because we sequence overlapping sections only once, except for a short overlap.  
 The true left end of clone RP11-295B17 is at 104579 in this sequence. The true right end of clone RP11-261P24 is at 2000 in this sequence.

## FEATURES

Source	Location/Qualifiers
	1..106578
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="13"
	/map="q31.3-32.3"
	/clone="RP11-111L24"
	/clone_lib="RPC1-11.1"
	79533
misc_feature	/note="Tandem repeat. Forced join. Gap size estimated to be approximately 800bp by restriction digest data"
BASE COUNT	28669 a 25979 c 25140 g 26790 t
ORIGIN	
Query Match	31.3%; Score 241; DB 9; Length 106578;
Best Local Similarity	100.0%; Pred. No. 1.le-47;
Matches 241; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY 219	AAGCACTGGCGGCTCACCTGTGGAGACAGCGAGGCGCTTGAGCGCCTGGAGAAATGGA 278
Db 37298	AGCACTGGCGGCTCACCTGTGGAGACAGCGAGGCGCTTGAGCGCCTGGAGAAATGGA 37357
QY 279	ATCAAGAGCTCCGGCGGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAGGTTGT 338
Db 37358	ATCAAGAGCTCCGGCGGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAGGTTGT 37417
QY 339	TACCTACCGCTCAACACCTTCTCTGCGGCACTGCACGGCTCATGACATCAAGCAG 398
Db 37418	TACCTACCGCTCAACACCTTCTCTGCGGCACTGCACGGCTCATGACATCAAGCAG 37477
QY 399	GTCTGGAGCGGCTGTGCAACACCAACCGAGCGGCGGCGGCTGAGGAGCTCCGGA 458
Db 37478	GTCTGGAGCGGCTGTGCAACACCAACCGGCGGCGGCGGCTGAGGAGCTCCGGA 37537
QY 459	G 459
Db 37538	G 37538
RESULT 8	
AC094777	
LOCUS	AC094777 224310 bp DNA linear HTG 24-AUG-2002
DEFINITION	Rattus norvegicus clone CH230-4N8, *** SEQUENCING IN PROGRESS ***
ACCESSION	AC094777
VERSION	AC094777.6 GI:22164997

## KEYWORDS

HTG; HTGS\_PHASE1.

## SOURCE

Norway rat.

## ORGANISM

Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

## REFERENCE

1 (bases 1 to 224310)

## AUTHORS

Muzny,D,Marie, Metzker,M, Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biviano, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flag, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Gant, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensu, H., Loulseg, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwako, O., Okwuonu, G., Otlorunpung, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smaj, D., Sneed, A., Sodergren, E., Song, X., Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

## TITLE

Direct Submission

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 224310)

## AUTHORS

Worley, K.C.

## TITLE

Direct Submission

## JOURNAL

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

## REFERENCE

3 (bases 1 to 224310)

## AUTHORS

Rat Genome Sequencing Consortium.

## TITLE

Direct Submission

## JOURNAL

Submitted (24-AUG-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Aug 9, 2002 this sequence version replaced gi:22094284.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GBIZ

Center clone name: CH230-4N8

----- Summary Statistics

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 172428 bases at least Q40

Consensus quality: 179331 bases at least Q30

Consensus quality: 184096 bases at least Q20

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 65 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1011: contig of 1011 bp in length

\* 1012 1111: gap of unknown length

\* 1112 2458: contig of 1347 bp in length

\* 2459 2558: gap of unknown length

\* 2559 3836: contig of 1278 bp in length

\* 3837 3936: gap of unknown length

\* 3937 5467: contig of 1531 bp in length

\* 5468 5567: gap of unknown length

\* 5568 6863: contig of 1296 bp in length

\* 6864 6963: gap of unknown length

\* 6964 8013: contig of 1050 bp in length

\* 8014 8113: gap of unknown length

\* 8114 9431: contig of 1318 bp in length

\* 9432 9531: gap of unknown length

\* 9532 10970: contig of 1439 bp in length

\* 10971 11070: gap of unknown length

\* 11071 12178: contig of 1108 bp in length

\* 12179 12278: gap of unknown length

\* 12279 13604: contig of 1326 bp in length

\* 13605 13704: gap of unknown length

\* 13705 15160: contig of 1456 bp in length

\* 15161 15260: gap of unknown length

\* 15261 16520: contig of 1260 bp in length

\* 16521 16620: gap of unknown length

\* 16621 18279: contig of 1659 bp in length

\* 18280 18379: gap of unknown length

\* 18380 19562: contig of 1183 bp in length

\* 19563 19662: gap of unknown length

\* 19663 20728: contig of 1066 bp in length

\* 20729 20828: gap of unknown length

\* 20829 22059: contig of 1231 bp in length

\* 22060 22159: gap of unknown length

\* 22160 24977: contig of 2818 bp in length

\* 24978 25077: gap of unknown length

\* 25078 26432: contig of 1355 bp in length

\* 26433 26532: gap of unknown length

\* 26533 27664: contig of 1132 bp in length

\* 27665 27764: gap of unknown length

\* 27765 29834: contig of 2070 bp in length

\* 29835 29934: gap of unknown length

\* 29935 32076: contig of 2142 bp in length

\* 32077 32176: gap of unknown length

\* 32177 33953: contig of 1777 bp in length

\* 33954 34053: gap of unknown length

\* 34054 36346: contig of 2293 bp in length

\* 36347 36446: gap of unknown length

\* 36447 38429: contig of 1983 bp in length

\* 38430 38529: gap of unknown length

\* 38530 41113: contig of 2584 bp in length

\* 41114 41213: gap of unknown length

\* 41214 43096: contig of 1883 bp in length

\* 43097 43196: gap of unknown length

\* 43197 44735: contig of 1539 bp in length

\* 44736 44835: gap of unknown length

\* 44836 46649: contig of 1814 bp in length

\* 46650 46749: gap of unknown length

\* 46750 48470: contig of 1721 bp in length

\* 48471 48570: gap of unknown length

\* 48571 50495: contig of 1925 bp in length

\* 50496 50595: gap of unknown length

\* 50596 52568: contig of 1973 bp in length

\* 52569 52668: gap of unknown length

\* 52669 54428: contig of 1760 bp in length

\* 54429 54528: gap of unknown length

\* 54529 58301: contig of 3773 bp in length

\* 58302 58401: gap of unknown length

\* 58402 62471: contig of 4070 bp in length

\* 62472 62571: gap of unknown length

\* 62572 64513: contig of 1942 bp in length

\* 64514 64613: gap of unknown length

\* 64614 66942: contig of 2329 bp in length

\* 66943 67042: gap of unknown length

\* 67043 69741: contig of 2699 bp in length

\* 69742 69841: gap of unknown length

\* 69842 72553: contig of 2712 bp in length

\* 72554 72653: gap of unknown length

\* 72654 74402: contig of 1749 bp in length

\* 74403 74502: gap of unknown length

\* 74503 77297: contig of 2795 bp in length

\* 77298 77397: gap of unknown length

\* 77398 81383: contig of 3986 bp in length

\* 81384 81484: gap of unknown length

\* 81484 83907: contig of 2424 bp in length

\* 83908 84007: gap of unknown length

\* 84008 86993: contig of 2986 bp in length

\* 86994 87093: gap of unknown length

\* 87094 89399: contig of 2306 bp in length

\* 89400 89499: gap of unknown length

\* 89500 93060: contig of 3561 bp in length

\* 93061 93160: gap of unknown length

\* 93161 97688: contig of 4528 bp in length

\* 97689 97788: gap of unknown length

\* 97789 102376: contig of 4588 bp in length

\* 102377 102476: gap of unknown length

\* 102477 106253: contig of 3777 bp in length

\* 106254 106353: gap of unknown length

\* 106354 109283: contig of 2930 bp in length

\* 109284 109383: gap of unknown length

\* 109384 112196: contig of 2813 bp in length

\* 112197 112296: gap of unknown length

Query Match 26.6%; Score 204.2; DB 2; Length 224310;

Best Local Similarity 90.5%; Pred. No. 9.8e-39;

Matches 218; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 219 AAGCACCTGGCGGCTCACCTGTGGAAGCACACGAGGCCCTTGAGGCCCTGGAGATGGA 278

DB 79335 AAGCACCTGGCGGCCACCTCTGGAAGCACACGAGGCCCTTGAGGCCCTGGAGACCTCC 79394

QY 279 ATCAAGAGCTCCGGCGGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAGGTGTGT 338

DB 79395 ATCAAGGGCTCCGGCGGCTGGAACACTTCTGCCGAGCTTCGAGTTGAGAAGGTGTGT 79454

QY 339 TACCTTACCGCTCAACACCTTCTCTCGGGCCACTGCACCGCTCATGCACATACAGCAG 398

DB 79455 TACCTTACCGCTCAACACCTTCTCTCGGGCCACTGCACCGCTCATGCACATACAGCAG 79514

QY 399 GTCTCTGGAGCGGCTGTGCAACACACCGCGGAGCGGAGCGGAGCTTTCAGGAGCTCCCA 458

DB 79515 GTCTCTGGAGCGGCTGTGCAACACATCACCTCCGACCATTCGAGCTTCAGGAGCTCCCA 79574

QY 459 G 459

DB 79575 G 79575



RESULT 9

AC122885

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AC122885 221502 bp DNA linear HTG 23-JUN-2002  
Mus musculus chromosome UNK clone RP23-90H14, WORKING DRAFT  
SEQUENCE, 8 unordered pieces.

AC122885  
HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
house mouse.

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 221502)  
McPherson,J.D. and Waterston,R.H.  
The sequence of Mus musculus clone  
Unpublished

2 (bases 1 to 221502)  
McPherson,J.D. and Waterston,R.H.  
Direct Submission  
Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA

3 (bases 1 to 221502)  
McPherson,J.D. and Waterston,R.H.  
Direct Submission  
Submitted (23-JUN-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA

On Jun 23, 2002 this sequence version replaced gi:21218511.  
----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)

----- Project Information -----  
Center project name: M\_BA0090H14  
----- Summary Statistics -----  
Sequencing vector: M13; 0%

Chemistry: Dye-primer; 100%  
Chemistry: Dye-terminator; Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319

Consensus quality: 216976 bases at least Q40  
Consensus quality: 217754 bases at least Q30  
Consensus quality: 218341 bases at least Q20

Insert size: 190000; agarose-gel  
Quality coverage: 15.60 in Q20 bases; agarose-gel  
Quality coverage: 12.41 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 8441: contig of 8441 bp in length  
\* 8442 8541: gap of unknown length  
\* 8542 27277: contig of 18736 bp in length

\* 27278 27377: gap of unknown length  
\* 27378 42339: contig of 14962 bp in length  
\* 42340 42439: gap of unknown length

\* 42440 71538: contig of 29099 bp in length  
\* 71539 71639: gap of unknown length  
\* 71639 99587: contig of 27949 bp in length

\* 99588 134048: contig of 34361 bp in length  
\* 134049 221349: contig of 87201 bp in length  
\* 134149

FEATURES

Source

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

Fax: (650) 320-5801  
Email: olivier@hgsc.stanford.edu  
Primer A: AGAGGATGCTTATCTCCCTCCG  
Primer B: ACGGTAGGATGATCAAGATGGAG  
STS size: 290  
PCR Profile:  
Initial incubation: 95 degrees C for 10 minutes  
Denaturation: 94 degrees C for 30 seconds  
Annealing: 60 degrees C for 30 seconds  
Polymerization: 72 degrees C for 23 seconds  
PCR Cycles: 30  
Thermal Cycler: Perkin Elmer 9700  
Protocol:  
Template: 25 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
Amplifrag Gold Polymerase: 0.07 units/uL  
Total Vol: 5 uL  
Buffer: MgCl2: 2.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 8.3  
Finished human sequence in NCBI. STSs designed and developed at the Stanford Human Genome Center.  
Location/Qualifiers  
1.393  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="13"  
/clone\_lib="Human"  
12..301  
12..34  
primer\_bind complement(279..301)  
primer\_bind 80 a 105 c 93 g 115 t  
ORIGIN  
Query Match 21.1%; Score 162.4; DB 11; Length 393;  
Best Local Similarity 95.9%; Pred. No. 8.7e-29;  
Matches 188; Conservative 0; Mismatches 6; Indels 2; Gaps 2;  
QY 409 GGCTGTGCAAAACA-CCACCCGCGAGCCGACTTCAGGACTGCCGAGCCGCTTG 467  
|||||  
Db 393 GGCTGTGCAAAACACCACCAGCCGAGCCGACTTCAGGACTGCCGAGCCGCTTG 334  
QY 468 GCAGAGATCACGGAGATGGTGGCACAGC-TCCACGGTACGATGATCAAGATGGAGAAATT 526  
|||||  
Db 333 GCAGAGATCACGGAGATGGTGGCACAGCTTCCACGGTACGATGATCAAGATGGAGAAATT 274  
QY 527 CCAGAGCTGCAGGAACTCAAGAAATTTGATGGCATTTGATGATTTGTTCCCGG 586  
Db 273 CCAGAGCTGCAGGAACTCAAGAAATTTGATGGCATTTGATGATTTGTTCCCGG 214  
QY 587 AAGGGAGTTTCATCCGT 602  
|||||  
Db 213 AAGGGTAAGCAGCAGT 198  
RESULT 11  
AC022669  
LOCUS AC022669 152053 bp DNA linear HTG 24-AUG-2002  
DEFINITION Homo sapiens chromosome 13 clone RP11-96B23 map 13, WORKING DRAFT  
SEQUENCE, 33 unordered pieces.  
ACCESSION AC022669  
VERSION AC022669.4 GI:9966291  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 152053)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Homo sapiens chromosome 13, clone RP11-96B23  
Unpublished  
2 (bases 1 to 152053)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,  
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,  
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,  
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A., and Zody,M.  
Direct Submission  
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 152053)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,  
Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,  
Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K.,  
Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,  
Iliev,I., Johnson,R., Jones,C., Karatas,A., Larocque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
Maclean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,  
McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,  
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,  
Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,  
Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 4, 2000 this sequence version replaced gi:7229803.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L4938  
Center clone name: 96\_B\_23  
----- Summary Statistics  
Sequencing vector: M13; M7815; 93% of reads  
Sequencing vector: Plasmid; n/a; %-0.f% of reads  
7.39091718610864Chemistry: Dye-terminator Big Dye; 100% of  
reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 136451 bases at least Q40  
Consensus quality: 143612 bases at least Q30  
Consensus quality: 146365 bases at least Q20  
Insert size: 158000; agarose-fp  
Insert size: 148853; sum-of-contents  
Quality coverage: 3.4 in Q20 bases; agarose-fp

Quality cov.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 33 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 784: contig of 784 bp in length  
\* 785 884: gap of 100 bp  
\* 885 1956: contig of 1072 bp in length  
\* 1957 2056: gap of 100 bp  
\* 2057 3113: contig of 1057 bp in length  
\* 3114 3213: gap of 100 bp  
\* 3214 4817: contig of 1604 bp in length  
\* 4818 4917: gap of 100 bp  
\* 4918 6582: contig of 1665 bp in length  
\* 6583 6682: gap of 100 bp  
\* 6683 8059: contig of 1377 bp in length  
\* 8060 8159: gap of 100 bp  
\* 8160 22369: contig of 14210 bp in length  
\* 22370 22469: gap of 100 bp  
\* 22470 24974: contig of 2505 bp in length  
\* 24975 25074: gap of 100 bp  
\* 25075 27206: contig of 2132 bp in length  
\* 27207 27306: gap of 100 bp  
\* 27307 29420: contig of 2114 bp in length  
\* 29421 29520: gap of 100 bp  
\* 29521 32000: contig of 2480 bp in length  
\* 32001 32100: gap of 100 bp  
\* 32101 36330: contig of 4230 bp in length  
\* 36331 36430: gap of 100 bp  
\* 36431 39067: contig of 2637 bp in length  
\* 39068 39167: gap of 100 bp  
\* 39168 42564: contig of 3397 bp in length  
\* 42565 42664: gap of 100 bp  
\* 42665 46664: contig of 4000 bp in length  
\* 46665 46764: gap of 100 bp  
\* 46765 51047: contig of 4283 bp in length  
\* 51048 51147: gap of 100 bp  
\* 51148 55269: contig of 4122 bp in length  
\* 55270 55369: gap of 100 bp  
\* 55370 58831: contig of 3462 bp in length  
\* 58832 58931: gap of 100 bp  
\* 58932 64077: contig of 5146 bp in length  
\* 64078 64177: gap of 100 bp  
\* 64178 68808: contig of 4631 bp in length  
\* 68809 68908: gap of 100 bp  
\* 68909 74170: contig of 5262 bp in length  
\* 74171 74270: gap of 100 bp  
\* 74271 79296: contig of 5026 bp in length  
\* 79297 79396: gap of 100 bp  
\* 79397 84724: contig of 5328 bp in length  
\* 84725 84824: gap of 100 bp  
\* 84825 89713: contig of 4889 bp in length  
\* 89714 89813: gap of 100 bp  
\* 89814 94721: contig of 4908 bp in length  
\* 94722 94821: gap of 100 bp  
\* 94822 100196: contig of 5375 bp in length  
\* 100197 100296: gap of 100 bp  
\* 100297 107015: contig of 6719 bp in length  
\* 107016 107115: gap of 100 bp  
\* 107116 114507: contig of 7392 bp in length  
\* 114508 114607: gap of 100 bp  
\* 114608 122671: contig of 8064 bp in length  
\* 122672 122771: gap of 100 bp  
\* 122772 128052: contig of 5281 bp in length  
\* 128053 128152: gap of 100 bp  
\* 128153 138583: contig of 10431 bp in length  
\* 138584 138683: gap of 100 bp  
\* 138684 150248: contig of 11565 bp in length  
\* 150249 150348: gap of 100 bp

\* 150349 152053: contig of 1705 bp in length.  
FEATURES  
Source  
Location/Qualifiers  
1. .152053  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/map="13"  
/clone="RP11-96B23"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. .784  
/note="assembly\_fragment"  
clone\_end:Sp6  
vector\_side:left"  
885. .1956  
/note="assembly\_fragment"  
2057. 3113  
/note="assembly\_fragment"  
3214. .4817  
/note="assembly\_fragment"  
4918. .6582  
/note="assembly\_fragment"  
6683. 8059  
/note="assembly\_fragment"  
8160. .22369  
/note="assembly\_fragment"  
22470. .24974  
/note="assembly\_fragment"  
25075. 27206  
/note="assembly\_fragment"  
27307. .29420  
/note="assembly\_fragment"  
29521. 32000  
/note="assembly\_fragment"  
32101. 36330  
/note="assembly\_fragment"  
36431. 39067  
/note="assembly\_fragment"  
39168. 42564  
/note="assembly\_fragment"  
42665. 46664  
/note="assembly\_fragment"  
46765. 51047  
/note="assembly\_fragment"  
Query Match 17.8%; Score 136.8; DB 2; Length 152053;  
Best Local Similarity 98.6%; Pred. No. 2.1e-22;  
Matches 138; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 8 GTGGTTTCAGAGCACAGTCAGCAAGAGAGGACGCCATGCCGGAAGTCTCAT 67  
|||||  
Db 103907 GTGGTTTCAGAGCACAGTCAGCAAGAGAGGACGCCATGCCGGAAGTCTCAT 103966  
QY 68 ATTCCCGAATTTTGACCTTTGCACAAATTTTCATCTAATTTTCTCAAGAAATTCAGCA 127  
|||||  
Db 103967 ATTCCCGAATTTTGACCTTTGCACAAATTTTCATCTAATTTTCTCAAGAAATTCAGCA 104026  
QY 128 ACGACTTGCCCTGTGGGAAG 147  
|||||  
Db 104027 ACGACTTGCCCTGTGGTGA 104046  
RESULT 12  
AX261839/c  
LOCUS AX261839 426 bp DNA linear PAT 26-OCT-2001  
DEFINITION Sequence 1490 from Patent WO0173027.  
ACCESSION AX261839  
VERSION AX261839.1 GI:16510806  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1

AUTHORS Meagher, M.J., Xu, J. and King, G.E.  
 TITLE Compositions and methods for therapy and diagnosis of colon cancer  
 JOURNAL Patent: WO 0173027-A 1490 04-OCT-2001;  
 CORIXA CORPORATION (US)  
 FEATURES Location/Qualifiers  
 source 1..426  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 89 a 109 c 109 g 119 t  
 ORIGIN

Query Match 15.7%; Score 120.8; DB 6; Length 426;  
 Best Local Similarity 60.2%; Pred. No. 1.1e-18;  
 Matches 219; Conservative 0; Mismatches 142; Indels 3; Gaps 1;

QY 1 TCACCTCGTGGTTTCAGGCACAGTCAGCAAGAGAGAGCCATGCGCGGAGACACTGCAAAA 60  
 DB 365 TTACCGTGTGGTTCCGCAGCGAGTGGTGAAGGAGGAGCCATGCGCTGCGACTCTGATGA 306  
 QY 61 GTCTCATATTCGCCGAATTTTGAACCTTTTCACAAATTTTCATACTAATTTTCACAGGAAA 120  
 DB 305 CGTGCTCTCTTCACACATCGATCCATCTATGAGTTCACAGAGGCTTCCTGCGCGAGG 246  
 QY 121 TTGAGCAGCACTGCCCTGTGGGAAGCGCGCTCAAAATGCCAAATTCAGAG---ATTACC 177  
 DB 245 TGGAGCAGAGGCTGGCACTCTGGGAAGGCGCTCCAAAGCCACACAAAAGCGAGTCATC 186  
 QY 178 AAGATCGCCATGTCATGCTGAAGACATTCAGGCGCATGAGCACCTGGCGGGTCACCC 237  
 DB 185 ACGAATCGGGACATCTCGCTCAGAGATCGCGCAGTTAAAGGAGTTTACCACTACT 126  
 QY 238 TGTGAAGCAGCAGCGAGGCGCTTGGAGGCGCTTGAAGTGAATGAATCAAGAGCTCCCGCGCGC 297  
 DB 125 TCCAAAGCATGACGAGATCCTTAACAGACTGGAAGAGCTACCAACGCTGTAGAAGT 56  
 QY 298 TGGAGAACTTTCGACAGACTTTGAGCTGCAGAGAGGTGTTTACCTACCGCTCAACACT 357  
 DB 65 TGGAGCAGTGTACAGAGGAGTTTGAGCTGCAGAGGCTGTGCTACTTGCCTCTCAACACT 6  
 QY 358 TCCT 361  
 DB 5 TCCT 2

RESULT 13  
 AC101725/c  
 LOCUS AC101725 190588 bp DNA linear HTG 21-AUG-2002  
 DEFINITION Mus musculus clone RP23-321M11, WORKING DRAFT SEQUENCE, 28  
 unordered pieces.  
 ACCESSION AC101725  
 VERSION AC101725.2 GI:22381046  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE house mouse  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 190588)  
 Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus, clone RP23-321M11  
 Unpublished  
 2 (bases 1 to 190588)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhalter, B.,  
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,  
 Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Illiev, I., Johnson, R.,  
 Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,  
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,

Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 190588)  
 Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
 Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B.,  
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,  
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
 Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,  
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,  
 Horton, L., Hulme, W., Illiev, I., Johnson, R., Jones, C., Kamat, A.,  
 Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,  
 Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,  
 McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,  
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
 Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P.,  
 Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,  
 Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
 Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 21, 2002 this sequence version replaced gi:17060500.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence.submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L17278  
 Center clone name: 321\_M11  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 182743 bases at least Q40  
 Consensus quality: 185393 bases at least Q30  
 Consensus quality: 186552 bases at least Q20  
 Insert size: 206000; agarose-fp  
 Insert size: 187888; sum-of-ctnigs  
 Quality coverage: 5.7 in Q20 bases; agarose-fp  
 Quality coverage: 6.3 in Q20 bases; sum-of-ctnigs  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 28 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 698: contig of 698 bp in length  
 \* 699 798: gap of 100 bp  
 \* 799 1437: contig of 639 bp in length  
 \* 1438 1537: gap of 100 bp  
 \* 1538 2255: contig of 718 bp in length  
 \* 2256 2355: gap of 100 bp

```
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCES
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 Contact: submissions@wustl.wustl.edu  
 ----- Project Information -----  
 Center project name: M.BA090H14  
 ----- Summary Statistics -----  
 Sequencing vector: M13; 0%  
 Sequencing vector: plasmid; 100%  
 Chemistry: Dye-primer ET; 0% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 216976 bases at least Q40  
 Consensus quality: 217754 bases at least Q30  
 Consensus quality: 218341 bases at least Q20  
 Insert size: 190000; agarose-fp  
 Quality coverage: 15.60 in Q20 bases; agarose-fp  
 Quality coverage: 12.41 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

* 1 8441: contig of 8441 bp in length
* 8442 8541: gap of unknown length
* 8542 27277: contig of 18736 bp in length
* 27278 27377: gap of unknown length
* 27378 42339: contig of 14962 bp in length
* 42340 42439: gap of unknown length
* 42440 71538: contig of 29099 bp in length
* 71539 99587: gap of unknown length
* 99588 99687: contig of 27949 bp in length
* 99688 134048: gap of unknown length
* 134149 221449: contig of 34361 bp in length
* 221450 221449: contig of 87201 bp in length
* 221350 221449: gap of unknown length
* 221450 221502: contig of 53 bp in length.

```

## FEATURES

Source

```

1. .221502
  /organism="Mus musculus"
  /db_xref="taxon:10090"
  /chromosome="UNK"
  /clone="RP23-90H14"
  misc_feature 1. 8441
    /note="assembly_name:Contig59"
  misc_feature 8542. .27277
    /note="assembly_name:Contig60"
  misc_feature 27378. .42339
    /note="assembly_name:Contig61"
  misc_feature 42440. 71538
    /note="assembly_name:Contig62"
  misc_feature 71639. .99587
    /note="assembly_name:Contig63"
  misc_feature 99688. .134048
    /note="assembly_name:Contig64"
  misc_feature 134149. .221349
    /note="assembly_name:Contig65"
  misc_feature 221450. .221502
    /note="assembly_name:Contig19"
  BASE COUNT 61405 a 54522 c 50450 g 54373 t 752 others
  ORIGIN

```

Query Match 15.5%; Score 119.2; DB 2; Length 221502;  
 Best Local Similarity 90.7%; Pred. No. 3.9e-18;  
 Matches 127; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

QY      8 GTGCTTTGAGCAGCAGTGCAGCAAGAGAGCGCCATCGCGAAGCACTGAAAGTCTCAT 67
        |||||
Db 111066 GTGCTTTGAGCAGCAGTGCAGCAAGAGAGCGCTCCATCGCGAAGCCCTGAAGAGTCTCAT 111007
        |||||

QY      68 ATTCGCCGAATTTTGAACCTTTGACAAATTTCTACTACTATTTTCTCAAGGAAATTTGACGA 127
        |||||
Db 111006 ATTCGCCGAATTTTGAACCTTTGACAAATTTCTACTACTATTTTCTCAAGGAAATTTGACGA 110947
        |||||

QY      128 ACGACTTGCCTGTGGGAAG 147
        - |||||
Db 110946 ACGACTTGCCTGTGGTGAG 110927
        |||||

```

## RESULT 15

AC118779

LOCUS

DEFINITION

\*\*\*, 61 unordered pieces.

AC118779

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa;

Mammalia; Eutheria;

Rodentia; Sciurognathi;

Muridae; Murinae;

Rattus.

1 (bases 1 to 149417)

REFERENCE

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Aisbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,

Barbata,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,

Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,N.P.,

Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,

Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,

Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,

Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,

Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,

Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,

Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,

Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,

Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,

Honsi,F., Howard,S., Huber,J., Huiy,J., Hume,J., Jackson,L.E.,

Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,

Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,

Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,

Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulsged,H.,

Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,

Maheshwari,M., Mapua,P., Martin,K., Martindale,A., Martinez,E.,

Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,

Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,

Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,

Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okwuonu,G.,

Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,

Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,

Rives,M., Rojas,A., Rojibokan,I., Roife,M., Ruiz,S., Saverly,G.,

Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,

Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,

Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,

Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,

Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q.,

Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,

Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,

Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 149417)

REFERENCE

AUTHORS

Worley,K.C.

Direct Submission

TITLE

JOURNAL

REFERENCE

2 (bases 1 to 149417)

Worley,K.C.

Direct Submission

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Submitted (21-APR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 149417)  
Worley, K.C.

## Direct Submission

Submitted (18-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 14, 2002 this sequence version replaced gi:20258202.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GUWA

Center clone name: CH230-255J6

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 106401 bases at least Q40

Consensus quality: 113612 bases at least Q30

Consensus quality: 117819 bases at least Q20

-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: this is a 'working draft' sequence. It currently

\* consists of 61 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1

1127: contig of 1127 bp in length

1128: gap of unknown length

1228: contig of 1008 bp in length

2236: gap of unknown length

2336: contig of 1457 bp in length

3793: gap of unknown length

3893: contig of 1085 bp in length

4978: gap of unknown length

5078: contig of 1029 bp in length

6107: gap of unknown length

6207: contig of 1411 bp in length

7618: gap of unknown length

9306: contig of 1388 bp in length

9406: gap of unknown length

10633: contig of 1227 bp in length

10733: gap of unknown length

12159: contig of 1426 bp in length

12259: gap of unknown length

13281: contig of 1023 bp in length

13381: gap of unknown length

13382: contig of 1683 bp in length

15065: gap of unknown length

15165: contig of 2554 bp in length

17119: gap of unknown length

17819: contig of 1545 bp in length

19364: gap of unknown length

20527: contig of 1063 bp in length

20627: gap of unknown length

21864: contig of 1238 bp in length

21865: gap of unknown length

21965: contig of 1874 bp in length

23838: gap of unknown length

23939: contig of 1096 bp in length

25034: gap of unknown length

25135: contig of 2236 bp in length

27371: gap of unknown length

28692: contig of 1222 bp in length

28792: gap of unknown length

30075: contig of 1283 bp in length

30175: gap of unknown length

31470: contig of 1295 bp in length

31570: gap of unknown length

32895: contig of 1325 bp in length

32995: gap of unknown length

34801: contig of 1806 bp in length

34901: gap of unknown length

36662: contig of 1761 bp in length

36762: gap of unknown length

37997: contig of 1235 bp in length

38097: gap of unknown length

40561: contig of 2464 bp in length

40661: gap of unknown length

42524: contig of 1863 bp in length

42624: gap of unknown length

44247: contig of 1623 bp in length

44348: gap of unknown length

46510: contig of 2163 bp in length

46611: gap of unknown length

48669: contig of 2059 bp in length

48769: gap of unknown length

50218: contig of 1449 bp in length

50318: gap of unknown length

52454: contig of 2136 bp in length

52455: gap of unknown length

53956: contig of 1402 bp in length

54056: gap of unknown length

56174: contig of 2118 bp in length

56274: gap of unknown length

58482: contig of 2208 bp in length

58582: gap of unknown length

60511: contig of 1929 bp in length

60611: gap of unknown length

62512: contig of 1901 bp in length

62612: gap of unknown length

65352: contig of 2740 bp in length

65452: gap of unknown length

67770: contig of 2318 bp in length

67870: gap of unknown length

70795: contig of 2925 bp in length

70895: gap of unknown length

73397: contig of 2502 bp in length

73497: gap of unknown length

75528: contig of 2031 bp in length

75628: gap of unknown length

78573: contig of 2945 bp in length

78673: gap of unknown length

80983: contig of 2210 bp in length

80984: gap of unknown length

83541: contig of 2558 bp in length

83641: gap of unknown length

86220: contig of 2579 bp in length

86320: gap of unknown length

89057: contig of 2737 bp in length

89157: gap of unknown length

92362: contig of 3205 bp in length

92462: gap of unknown length

95949: contig of 3487 bp in length

96049: gap of unknown length

100396: contig of 4347 bp in length

100496: gap of unknown length

104325: contig of 3829 bp in length

104425: gap of unknown length

107324: contig of 2899 bp in length

107424: gap of unknown length

110542: contig of 3118 bp in length

110642: gap of unknown length

113543: contig of 2901 bp in length

Query Match 14.5%; Score 111.8; DB 2; Length 149417;  
Best Local Similarity 87.8%; Pred. No. 2.4e-16;

	Matches	122;	Conservative	0;	Mismatches	17;	Indels	0;	Gaps	0;
Qy	453	TGCGAGCGCTTTTGGCAGAGATCACGGAGATGGTGGCAGAGCTCCACGGTACGATGATC	512							
Db	28953	TCCTCAGCTGGCTGGCAGAGATCACAGAGATGGTGGCAGAGCTGCACGGTACCATGATC	29012							
Qy	513	AAGATGGAGAAATTTCCAGAAGCTGCACGAACCTCAAGAAAGATTTGATTGGCATTGACAAT	572							
Db	29013	AAGATGGAGAACTTCCAGAAGCTGCATGAGCTCAAGAAGGATCTGATTGSCATTGACAAT	29072							
Qy	573	CTTGTGGTTCCGGGAAGGG	591							
Db	29073	CTTGTGATCCAGGAAGGG	29091							

Search completed: December 12, 2002, 06:06:55  
Job time : 3622 secs



GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2002, 02:37:49 ; Search time 306 seconds

(without alignments)  
5659.434 Million cell updates/sec

Title: US-09-555-342b-1\_COPY\_1733\_2501

Perfect score: 769

Sequence: 1 tcaactcgtggttcagacg.....attgaggagacgaagacga 769

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	769	100.0	3187	23	AAS64760
2	769	100.0	3442	20	AAH79183
3	769	100.0	4687	21	AAA08582
4	762	99.1	3094	21	AAC98992
5	120.8	15.7	426	23	AAS58814
6	106	13.8	1718	22	AAS27016
7	105.2	13.7	716	22	AAS27440
8	98	12.7	717	24	ABL90828
9	97.2	12.6	1747	23	ABL03867

C	10	97.2	12.6	4150	23	ABL03866	Drosophila melanog
	11	90.2	11.7	544	22	AAH98023	Murine 7-transmemb
	12	83.6	10.9	665	22	AAH07867	Human CDNA clone (
	13	83.6	10.9	1296	24	ABK49898	Human CDNA encodin
	14	83.6	10.9	1821	22	AAH15014	Human CDNA sequenc
	15	83.6	10.9	3610	24	ABN59933	Novel human coding
C	16	72.8	9.5	462	24	ABK53702	Human eosinophil-m
	17	64.4	8.4	2686	22	AAH27013	CDNA encoding nove
	18	64.4	8.4	3321	22	AAH44718	Human full-length
	19	64	8.3	1091	22	AAH27281	CDNA encoding nove
	20	60.8	7.9	1860	22	AAF99965	Human colon carcin
	21	60.8	7.9	5923	22	AAH02052	DNA encoding molec
C	22	60.8	7.9	5923	22	AAH02052	DNA encoding molec
	23	60	7.8	60	24	ABN41498	Human spliced tran
	24	56.2	7.3	365	14	AAQ39923	Expressed Sequence
	25	56.2	7.3	365	14	AAQ59335	Human brain expres
	26	53.2	6.9	417	22	ABA08384	Human collybistin
C	27	46.4	6.0	454	23	ABV54945	Human prostate exp
	28	46.4	6.0	541	24	ABN64838	Human cancer relat
	29	44.4	5.8	342	21	AAF08796	Fusarium venenatum
	30	44	5.7	1609	23	ABL25953	Drosophila melanog
	31	44	5.7	3381	23	ABL25950	Drosophila melanog
	32	44	5.7	3609	23	ABL25952	Drosophila melanog
C	33	43.4	5.6	3786	23	AAH54228	Pseudomonas aerugi
	34	42.4	5.5	1730	22	AAH27017	CDNA encoding nove
	35	42.4	5.5	1730	22	AAH27441	CDNA encoding nove
	36	42.4	5.5	1730	22	AAH27441	CDNA encoding nove
C	37	42.4	5.5	1778	22	AAI60338	Human CDNA SEQ ID
	38	42.4	5.5	1816	22	AAI58552	Human polynucleoti
	39	42	5.5	3306	23	ABL12855	Drosophila melanog
	40	41.2	5.4	2159	15	AAQ67223	Mouse p55Nuc. Mus
	41	40.8	5.3	625	22	AAH07308	Human CDNA clone (
	42	40.8	5.3	1444	22	AAH41863	Human chondroadher
	43	40.8	5.3	1696	22	AAH16459	Human stomach canc
	44	40.8	5.3	1696	22	AAH16459	Human CDNA sequenc
	45	40.8	5.3	2172	23	AAH51456	Pseudomonas aerugi

ALIGNMENTS

RESULT 1  
AAS64760  
ID AAS64760 standard; CDNA; 3187 BP.  
XX  
AC AAS64760;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #564.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
(HYSE-) HYSEQ INC.  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
XX P-PSDB; ABG00573.  
PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX Claim 1; SEQ ID No 564; 103pp; English.  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 3187 BP; 810 A; 874 C; 876 G; 627 T; 0 other;  
Query Match 100.0%; Score 769; DB 23; Length 3187;  
Best Local Similarity 100.0%; Pred. No. 1.7e-201;  
Matches 769; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCACCTCGTGGTTTCAGACACAGTGGACAAAGAGAGACGCCATGCCGGAAGCACTGAAAA 60  
DB 1734 TCACCTCGTGGTTTCAGACACAGTGGACAAAGAGAGACGCCATGCCGGAAGCACTGAAAA 1793  
QY 61 GTCTCATATTCCTCCGAATTTTGAACCTTTTCACAAATTTTCATACATACTAATTTCTCAAGGAAA 120  
DB 1794 GTCTCATATTCCTCCGAATTTTGAACCTTTTCACAAATTTTCATACATACTAATTTCTCAAGGAAA 1853  
QY 121 TTGAGCAACGACTTGCCTGTGGAGGCGCTCAAAATGCCCAATTCAGAGATTACCAAA 180  
DB 1854 TTGAGCAACGACTTGCCTGTGGAGGCGCTCAAAATGCCCAATTCAGAGATTACCAAA 1913  
QY 181 GAATCGGCATGTCTGCTGAAGACATTCAGGCGATGAAGCACTGCGGCTCACCCTGT 240  
DB 1914 GAATCGGCATGTCTGCTGAAGACATTCAGGCGATGAAGCACTGCGGCTCACCCTGT 1973  
QY 241 GGAAGCACAGCGAGCGCTTGGAGGCGCTGGAGAATGGAATCAAGAGCTCCCGCGGCTGG 300  
DB 1974 GGAAGCACAGCGAGCGCTTGGAGGCGCTGGAGAATGGAATCAAGAGCTCCCGCGGCTGG 2033  
QY 301 AGAAGCTTCTGACAGACTTTGAGCTGCAGAGGTGTGTTTACCTACCGCTTCACACCTTCC 360  
DB 2034 AGAAGCTTCTGACAGACTTTGAGCTGCAGAGGTGTGTTTACCTACCGCTTCACACCTTCC 2093  
QY 361 TCCTGGCGGCATGCACCGGCTCATGCACTACAGCAGTCTCTGGAGCGGCTGTGCAAAAC 420  
DB 2094 TCCTGGCGGCATGCACCGGCTCATGCACTACAGCAGTCTCTGGAGCGGCTGTGCAAAAC 2153  
QY 421 ACCACCGCGGAGCCACGCGGCTTTCAGGAGCTGCGGAGCGGCTTGGCAGAGATCACGG 480  
DB 2154 ACCACCGCGGAGCCACGCGGCTTTCAGGAGCTGCGGAGCGGCTTGGCAGAGATCACGG 2213  
QY 481 AGATGGTGGCACAGCTTCCACGGTACGATGATCAAGATGGAGAAATTTCCAGAAGCTGCACG 540  
DB 2214 AGATGGTGGCACAGCTTCCACGGTACGATGATCAAGATGGAGAAATTTCCAGAAGCTGCACG 2273  
QY 541 AACTCAAGAAAGATTGATTGGCATTGACAATCTGTGTGGTTCGGGAAGGGAGTTTCAATCC 600  
DB 2274 AACTCAAGAAAGATTGATTGGCATTGACAATCTGTGTGGTTCGGGAAGGGAGTTTCAATCC 2333

QY 601 GTCTGGGAGCCTCAGCAAGCTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTCTGT 660  
DB 2334 GTCTGGGAGCCTCAGCAAGCTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTCTGT 2393  
QY 661 TCACGACGCTGCTGTATATACAGAGCGCGGGGCTCAGCGCTTCCCAATCAGTTTAAAGTCC 720  
DB 2394 TCACGACGCTGCTGTATATACAGAGCGCGGGGCTCAGCGCTTCCCAATCAGTTTAAAGTCC 2453  
QY 721 ACGGCGAGCTCCCGCTCTATGCGCATGACGATTTGAGGAGCGCAAGACGA 769  
DB 2454 ACGGCGAGCTCCCGCTCTATGCGCATGACGATTTGAGGAGCGCAAGACGA 2502  
RESULT 2  
AAX79183  
ID AAX79183 standard; DNA; 3442 BP.  
XX  
AC AAX79183;  
XX  
DT 17-AUG-1999 (first entry)  
XX Human chondrocyte-derived gene CDEP.  
XX  
DE Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;  
KW Dbl homology domain; pleckstrin homology domain; rheumatoid 'arthritis;  
KW drug; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9928458-A1.  
XX  
PD 10-JUN-1999.  
XX  
PF 27-NOV-1998; 98WO-JP05348.  
XX  
PR 27-NOV-1997; 97JP-0342060.  
XX  
PA (CHUS ) CHUGAI SEIYAKU KK.  
XX  
PI Kato Y, Kawamoto T, Koyano Y;  
XX  
DR WPI; 1999-371117/31.  
DR P-PSDB; AAY07482.  
XX  
PT Protein CDEP expressed in differentiated chondrocytes, and gene  
PT encoding it  
XX  
PS Claim 5; Fig 1; 59pp; Japanese.  
XX  
CC This sequence represents the coding region for a protein (CDEP) expressed  
CC in differentiated human foetal chondrocytes, which contains an ezrin-like  
CC domain, a Dbl homology (DH) domain and a pleckstrin homology (PH) domain.  
CC The nucleic acid or protein can be used in the investigation and  
CC treatment of cancers and arthritic diseases (including chronic rheumatoid  
CC arthritis), or for screening of candidate anticancer drugs.  
XX  
SQ Sequence 3442 BP; 864 A; 952 C; 927 G; 599 T; 0 other;  
Query Match 100.0%; Score 769; DB 20; Length 3442;  
Best Local Similarity 100.0%; Pred. No. 1.8e-201;  
Matches 769; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCACCTCGTGGTTTCAGAGCACAGTGGACAAAGAGAGACGCCATGCCGGAAGCACTGAAAA 60  
DB 1733 TCACCTCGTGGTTTCAGAGCACAGTGGACAAAGAGAGACGCCATGCCGGAAGCACTGAAAA 1792  
QY 61 GTCTCATATTCCTCCGAATTTTGAACCTTTGCACAAATTTTCATATTTCTCAAGGAAA 120  
DB 1793 GTCTCATATTCCTCCGAATTTTGAACCTTTGCACAAATTTTCATATTTCTCAAGGAAA 1852  
QY 121 TTGAGCAACGACTTGCCTGTGGAGGCGGCTCAAAATGCCCAATTCAGAGATTACCAAA 180



Db 2314 AACTCAAGAAAGATTGATGGCAATTCACAACTTCTGTGTCGGGAAGGAGTTCATCC 2373  
QY 601 GTCTGGCGAGCTTCAGCAAGCTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCCTGT 660  
Db 2374 GTCTGGCGAGCTTCAGCAAGCTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCCTGT 2433  
QY 661 TCAACGACGCTCTGCTATACAGAGCCGGGGGCTGACGGCCCTCCATCATAGTTTAAAGTCC 720  
Db 2434 TCAACGACGCTCTGCTATACAGAGCCGGGGGCTGACGGCCCTCCATCATAGTTTAAAGTCC 2493  
QY 721 ACGGCGAGCTCCGCTCTATGCGATGACGATTGAGGAGAGCGAAGACGA 769  
Db 2494 ACGGCGAGCTCCGCTCTATGCGATGACGATTGAGGAGAGCGAAGACGA 2542

RESULT 4  
ID AAC98992 standard; cDNA; 3094 BP.  
XX AAC98992;  
XX 09-MAR-2001 (first entry)  
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:220.  
XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;  
KW detection; diagnosis; identification; cytostatic; neuroprotective;  
KW neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;  
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;  
KW linkage analysis; tissue identification; tissue typing; forensic;  
KW neural; immune system; muscular; reproductive; gastrointestinal;  
KW pulmonary; cardiovascular; renal; proliferative; ss.  
XX Homo sapiens.  
OS  
XX WO200055320-A1.  
PN  
XX 21-SEP-2000.  
PD  
XX 08-MAR-2000; 2000WO-US05989.  
XX  
XX 12-MAR-1999; 99US-0124270.  
PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Ruben SM;  
PI  
XX WPI; 2000-579444/54.  
DR  
XX P-PSDB; AAB54227.  
XX  
XX New nucleic acid that is a pancreatic cancer antigen for preventing,  
PT treating, or ameliorating a medical condition, particular pancreatic  
PT cancer, or for use in assays for diagnosing a pathological condition -  
XX  
XX Claim 1; Page 664-665; 1379pp; English.  
XX  
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated  
CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
CC neuroprotective, neotropic, immunomodulatory, relaxant, contraceptive,  
CC gynaecological, cardiant and antiinflammatory activities, and can be used  
CC in gene therapy. The polynucleotide and proteins can be used for  
CC preventing, treating, or ameliorating a medical condition or in assays  
CC for diagnosing a pathological condition or a susceptibility to one in a  
CC subject. Binding partners to the proteins and the activity of the  
CC proteins can be identified. The pancreatic cancer antigens can be used to  
CC detect, treat or prevent pancreatic disorders, especially cancer.  
CC Agonists and antagonists to the antigens can be screened for. The  
CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
CC acid hybridisation probes that can be used in chromosome mapping, linkage  
CC analysis, tissue identification and/or typing and a variety of forensic  
CC and diagnostic methods. The proteins can be used to generate antibodies  
CC which are used to purify, detect and target the polypeptides, including

CC both in vivo and in vitro diagnostic and therapeutic methods. The  
CC proteins can be used to treat or prevent neural, immune system, muscular,  
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
CC sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 3094 BP; 736 A; 849 C; 755 G; 749 T; 5 other;  
Query Match 99.1%; Score 762; DB 21; Length 3094;  
Best Local Similarity 100.0%; Pred. No. 1.5e-199; Indels 0; Gaps 0;  
Matches 762; Conservative 0; Mismatches 0;  
QY 8 GTGGTTTCAGAGCAGTCAGCAAGAGGACGCCATGCGGGAAGCACTGAAAGTCTCAT 67  
Db 62 GTGGTTTCAGAGCAGTCAGCAAGAGGACGCCATGCGGGAAGCACTGAAAGTCTCAT 121  
QY 68 ATTCCCGAATTTTGAACCTTTGCACAAATTTTCACTAATTTTCTCAAGAAATTCGAC 127  
Db 122 ATTCCCGAATTTTGAACCTTTGCACAAATTTTCACTAATTTTCTCAAGAAATTCGAC 181  
QY 128 ACGACTTGCCCTGTGGGAAGCGGCTCAATGCCCAATCAGAGATTACCAAGAAATCGG 187  
Db 182 ACGACTTGCCCTGTGGGAAGCGGCTCAATGCCCAATCAGAGATTACCAAGAAATCGG 241  
QY 188 CGATGTCTATGCTGAAGAACATTCAGGGCATGAAGCACCTTGGGGCTTCACCTGTGAAGCA 247  
Db 242 CGATGTCTATGCTGAAGAACATTCAGGGCATGAAGCACCTTGGGGCTTCACCTGTGAAGCA 301  
QY 248 CAGCAGGCGCTTGGAGGCCCTGGAGAATGGAATCAAGAGCTCCCGGCGGCTGGAGAATTT 307  
Db 302 CAGCAGGCGCTTGGAGGCCCTGGAGAATGGAATCAAGAGCTCCCGGCGGCTGGAGAATTT 361  
QY 308 CTGCAGAGACTTTGAGCTGCAGAAAGTGTTACCTACCGCTCAACACCTTCTCTCTGGG 367  
Db 362 CTGCAGAGACTTTGAGCTGCAGAAAGTGTTACCTACCGCTCAACACCTTCTCTCTGGG 421  
QY 368 GCCACTGCACCGGCTCATGCACATAGAAGCAGGTCTCTGGAGCGGCTGTGCAAAACACACCC 427  
Db 422 GCCACTGCACCGGCTCATGCACATAGAAGCAGGTCTCTGGAGCGGCTGTGCAAAACACCC 481  
QY 428 GCCGAGCCACGCCGACTTCAGGAGCTGCGGAGCCGCTTTGGCAGAGATCACGAGATGTT 487  
Db 482 GCCGAGCCACGCCGACTTCAGGAGCTGCGGAGCCGCTTTGGCAGAGATCACGAGATGTT 541  
QY 488 GGCACAGCTCCACGGTACGATGATCAAGATGGAGATTTCCAGAGCTCGACGAACTCAA 547  
Db 542 GGCACAGCTCCACGGTACGATGATCAAGATGGAGATTTCCAGAGCTCGACGAACTCAA 601  
QY 548 GAAAGATTTGATTGGCATTGACAATCTTGTGTTCCGGGAAGGGAGTTTCATCCGCTGGG 607  
Db 602 GAAAGATTTGATTGGCATTGACAATCTTGTGTTCCGGGAAGGGAGTTTCATCCGCTGGG 661  
QY 608 CAGCCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCACGA 667  
Db 662 CAGCCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCACGA 721  
QY 668 CGTCTGCTATACAGAGCGCGGGGCTGACGGCTCCCAATCAGTTTAAAGTCCACGGGCA 727  
Db 722 CGTCTGCTATACAGAGCGCGGGGCTGACGGCTCCCAATCAGTTTAAAGTCCACGGGCA 781  
QY 728 GCTCCCGCTCTATGGCATGACGATTGAGGAGAGCGAAGACGA 769  
Db 782 GCTCCCGCTCTATGGCATGACGATTGAGGAGAGCGAAGACGA 823

RESULT 5  
AAS58814/c  
ID AAS58814 standard; cDNA; 426 BP.  
XX AAS58814;  
XX AAS58814;  
XX 13-FEB-2002 (first entry)  
XX



PR	05-SEP-2000;	2000US-0229513.	PR	17-NOV-2000;	2000US-0249217.	XX	
PR	06-SEP-2000;	2000US-0230437.	PR	17-NOV-2000;	2000US-0249218.	PR	
PR	06-SEP-2000;	2000US-0230438.	PR	17-NOV-2000;	2000US-0249244.	PR	
PR	08-SEP-2000;	2000US-0231242.	PR	17-NOV-2000;	2000US-0249245.	PR	
PR	08-SEP-2000;	2000US-0231243.	PR	17-NOV-2000;	2000US-0249264.	PR	
PR	08-SEP-2000;	2000US-0231244.	PR	17-NOV-2000;	2000US-0249265.	PR	
PR	08-SEP-2000;	2000US-0231413.	PR	17-NOV-2000;	2000US-0249297.	PR	
PR	08-SEP-2000;	2000US-0231414.	PR	17-NOV-2000;	2000US-0249299.	PR	
PR	08-SEP-2000;	2000US-0232080.	PR	17-NOV-2000;	2000US-0249300.	PR	
PR	12-SEP-2000;	2000US-0231968.	PR	01-DEC-2000;	2000US-0250160.	PR	
PR	14-SEP-2000;	2000US-0232397.	PR	01-DEC-2000;	2000US-0250391.	PR	
PR	14-SEP-2000;	2000US-0232398.	PR	05-DEC-2000;	2000US-0251030.	PR	
PR	14-SEP-2000;	2000US-0232399.	PR	05-DEC-2000;	2000US-0251988.	PR	
PR	14-SEP-2000;	2000US-0232400.	PR	05-DEC-2000;	2000US-0256719.	PR	
PR	14-SEP-2000;	2000US-0232401.	PR	06-DEC-2000;	2000US-0251479.	PR	
PR	14-SEP-2000;	2000US-0233063.	PR	08-DEC-2000;	2000US-0251856.	PR	
PR	14-SEP-2000;	2000US-0233064.	PR	08-DEC-2000;	2000US-0251868.	PR	
PR	14-SEP-2000;	2000US-0233065.	PR	08-DEC-2000;	2000US-0251869.	PR	
PR	21-SEP-2000;	2000US-0234223.	PR	08-DEC-2000;	2000US-0251989.	PR	
PR	21-SEP-2000;	2000US-0234274.	PR	08-DEC-2000;	2000US-0251990.	PR	
PR	25-SEP-2000;	2000US-0234997.	PR	11-DEC-2000;	2000US-0254097.	PR	
PR	25-SEP-2000;	2000US-0234998.	PR	05-JAN-2001;	2001US-02559678.	XX	
PR	26-SEP-2000;	2000US-0235484.	PR	(HUMA-) HUMAN GENOME SCI INC.		PA	
PR	27-SEP-2000;	2000US-0235834.	PR	Rosen CA, Barash SC, Ruben SM;		XX	
PR	29-SEP-2000;	2000US-0235836.	PR	WPI: 2001-465460/50.		PI	
PR	29-SEP-2000;	2000US-0236327.	PR	P-PSDB; AUL17099.		XX	
PR	29-SEP-2000;	2000US-0236367.	DR			XX	
PR	29-SEP-2000;	2000US-0236368.	DR			DR	
PR	29-SEP-2000;	2000US-0236369.	DR			DR	
PR	29-SEP-2000;	2000US-0236370.	XX			XX	
PR	02-OCT-2000;	2000US-0237037.	PT			PT	
PR	02-OCT-2000;	2000US-0237038.	PT			PT	
PR	02-OCT-2000;	2000US-0237039.	XX			XX	
PR	02-OCT-2000;	2000US-0237040.	XX			XX	
PR	13-OCT-2000;	2000US-0239935.	PR			CC	
PR	13-OCT-2000;	2000US-0239937.	PR			CC	
PR	20-OCT-2000;	2000US-0240960.	CC			CC	
PR	20-OCT-2000;	2000US-0241221.	CC			CC	
PR	20-OCT-2000;	2000US-0241785.	CC			CC	
PR	20-OCT-2000;	2000US-0241786.	CC			CC	
PR	20-OCT-2000;	2000US-0241787.	CC			CC	
PR	20-OCT-2000;	2000US-0241808.	CC			CC	
PR	20-OCT-2000;	2000US-0241809.	CC			CC	
PR	20-OCT-2000;	2000US-0241826.	CC			CC	
PR	01-NOV-2000;	2000US-0244617.	PR			CC	
PR	08-NOV-2000;	2000US-0246474.	PR			CC	
PR	08-NOV-2000;	2000US-0246475.	PR			CC	
PR	08-NOV-2000;	2000US-0246476.	PR			CC	
PR	08-NOV-2000;	2000US-0246477.	PR			CC	
PR	08-NOV-2000;	2000US-0246478.	PR			CC	
PR	08-NOV-2000;	2000US-0246523.	PR			CC	
PR	08-NOV-2000;	2000US-0246524.	PR			CC	
PR	08-NOV-2000;	2000US-0246525.	PR			CC	

||||| || ||||| || || ||||| ||||| ||||| || || || |||||  
Db 133 ATGTTTTTCTGTTCTCAGATATGTTGCTGTACAAAGCAAAAGGAGTTGCAGGGACCAAC 192  
QY 708 CAGTTTAAAGTCCAGCGCAGCTCCCGCTCTATGCATGACGATTGAGGAGCAAGAC 767  
Db 193 CACTTCCGGATCCGGGGCTCTCTCCCTCCAGGCAATGCTGGTGAAGAAAGTGATTAAC 252  
QY 768 GA 769  
Db 253 GA 254  
||  
RESULT 7  
ID AAS27440  
ID AAS27440 standard; cDNA; 716 BP.  
XX AC  
AC AAS27440;  
XX  
XX 07-NOV-2001 (first entry)  
DE  
XX  
XX cDNA encoding novel signal transduction pathway protein, Seq ID 475.  
XX  
XX Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;  
KW antineoplastic; anti-HIV; antibacterial; anti-inflammatory; cancer;  
KW immune system disorder; rheumatoid arthritis; inflammatory condition;  
KW organ transplant rejection; infection; hepatitis C; blood disorder;  
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;  
KW acquired immune deficiency syndrome.  
XX  
XX Homo sapiens.  
XX  
XX WO200154733-A1.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01312.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-MAR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241788.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.

PR	17-NOV-2000;	2000US-0249208.	
PR	17-NOV-2000;	2000US-0249209.	
PR	17-NOV-2000;	2000US-0249210.	
PR	17-NOV-2000;	2000US-0249211.	
PR	17-NOV-2000;	2000US-0249212.	
PR	17-NOV-2000;	2000US-0249213.	
PR	17-NOV-2000;	2000US-0249214.	
PR	17-NOV-2000;	2000US-0249215.	
PR	17-NOV-2000;	2000US-0249216.	
PR	17-NOV-2000;	2000US-0249217.	
PR	17-NOV-2000;	2000US-0249218.	
PR	17-NOV-2000;	2000US-0249244.	
PR	17-NOV-2000;	2000US-0249245.	
PR	17-NOV-2000;	2000US-0249264.	
PR	17-NOV-2000;	2000US-0249265.	
PR	17-NOV-2000;	2000US-0249297.	
PR	17-NOV-2000;	2000US-0249299.	
PR	17-NOV-2000;	2000US-0249300.	
PR	01-DEC-2000;	2000US-0250160.	
PR	01-DEC-2000;	2000US-0250391.	
PR	05-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251988.	
PR	05-DEC-2000;	2000US-0256719.	
PR	06-DEC-2000;	2000US-0251479.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251868.	
PR	08-DEC-2000;	2000US-0251869.	
PR	08-DEC-2000;	2000US-0251989.	
PR	08-DEC-2000;	2000US-0251990.	
PR	11-DEC-2000;	2000US-0254097.	
PR	05-JAN-2001;	2001US-0259678.	
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Rosen CA, Barash SC, Ruben SM;		
XX			
DR	WPI; 2001-465460/50.		
DR	P-PSDB; AAU17523.		
XX			
PT	Novel polypeptides useful for diagnosing, treating, preventing and/or		
PT	prognosing disorders related to the proteins, including cancers, immune		
PT	disorders and neuronal disorders		
XX			
PS	Claim 1; SEQ ID No 475; 880pp; English.		
XX			
CC	The invention relates to novel isolated polypeptides (I), and		
CC	polynucleotides (II). (I), (II) and the antibody to (I) are useful for		
CC	diagnosing, preventing and treating diseases including immune system		
CC	disorders (e.g. congenital and acquired immunodeficiencies, autoimmune		
CC	disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ		
CC	transplant rejections and graft versus host disease, infectious diseases		
CC	(e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and		
CC	other blood-related disorders (sickle cell anaemia), myeloproliferative		
CC	disorders, primary haematopoietic disorders, hyperproliferative		
CC	disorders (e.g. Gaucher's disease and cancer), neurodegenerative		
CC	disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal		
CC	abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal		
CC	disorders (e.g. glomerulonephritis), cardiovascular disorders		
CC	(e.g. arrhythmia), respiratory disorders, dermatological disorders, in		
CC	wound healing, epithelial cell proliferation, endocrine disorders (e.g.		
CC	Addison's disease), reproductive system disorders, gastrointestinal		
CC	disorder (inflammatory disorders), liver disorders (cirrhosis),		
CC	as stimulators of B-cell responsiveness to pathogens, activators of		
CC	T-cells, to induce higher affinity antibodies, and as a means to induce		
CC	tumour proliferation in pathologies e.g. acquired immune deficiency		
CC	syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction		
CC	pathway protein coding sequences and PCR primers of the invention.		
XX			
Query Match	13.7%;	Score 105.2;	DB 22; Length 716;
Best Local Similarity	64.0%;	Pred. No. 6.9e-19;	
Matches 155;	Conservative	2; Mismatches 85;	Indels 0; Gaps 0;
QY	528	CAGAAGCTGCACGAACCAAGAAATTGATGGCATTTGACAAATCTTGTGTTCCGGGA	587
DB	3	CAGAAGCTWACGGAGCTCCKWCGGACCTGGTGGCATAGAGAACCTCATTTGCTCCTGGC	62
QY	588	AGGAGTTTCATCCGCTCTGGGAGCTCAGCAAGTCTTCGGGAAAGGGGCTCCAGCAGCGC	647
DB	63	AGGAGTTTCATCCGCTGAGGGCTGCTTCAAGAGTTCACCAAGAGGGGCTCCAGCAGAGG	122
QY	648	ATGTTCTTCCTGTTCAAGCAGCTCTGCTATACAGGACCGGGGCTGACGGCTCCAAAT	707
DB	123	ATGTTTCTTCTGTTCTCAGATATGTTGCTGTACACAAGAGAGTTGACGGGACCAAGC	182
QY	708	CAGTTTAAAGTCCACGGGACCTCCCGCTCTATGTCATGACGATTTGAGGAGCGCAACAC	767
DB	183	CAC TTCGGATCCGGGGCTCTTCCCTCCAGGCATGCTGTTGGTGAAGAAGTGATAC	242
QY	768	GA 769	
DB	243	GA 244	
RESULT 8			
ABL90828			
ID	ABL90828	standard; cDNA; 717 BP.	
XX	AC	ABL90828;	
XX	DT	24-MAY-2002 (first entry)	
XX	DE	Human polynucleotide.SEQ ID NO 1390.	
XX	KW	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;	
KW	KW	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;	
KW	KW	vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;	
KW	KW	cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;	
XX	KW	neurological disease; infection; human; secreted protein; gene; ss.	
OS	XX	Homo sapiens.	
XX	PN	WO200190304-A2.	
XX	PD	29-NOV-2001.	
XX	PF	18-MAY-2001; 2001WO-US16450.	
XX	PR	19-MAY-2000; 2000US-205515P.	
XX	PA	(HUMA-) HUMAN GENOME SCI INC.	
XX	PI	Birse CE, Rosen CA;	
XX	DR	WPI; 2002-122018/16.	
XX	DR	P-PSDB; ABB90419.	
XX	PT	Novel 1405 isolated polypeptides, useful for diagnosis, treatment and	
PT	PT	prevention of neural, immune system, muscular, reproductive,	
PT	PT	gastrointestinal, pulmonary, cardiovascular, renal and proliferative	
PT	PT	disorders	
XX	PS	Claim 4; SEQ ID NO 1390; 2081pp + Sequence Listing; English.	
XX	CC	The invention relates to novel genes (ABL89449-ABL90853) and proteins	
CC	CC	(ABB89040-ABB90444) useful for preventing, treating or ameliorating	
CC	CC	medical conditions e.g. by protein or gene therapy. The genes are	
CC	CC	isolated from a range of human tissues disclosed in the specification.	
CC	CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful	
CC	CC	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast	
CC	CC	and ovarian cancer and other cancers of the adrenal gland, bone, bone	
CC	CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;	
CC	CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune	
CC	CC	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's	
CC	CC	disease, multiple sclerosis, rheumatoid arthritis and ulcerative	
CC	CC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;	



CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 717 BP; 170 A; 196 C; 203 G; 144 T; 4 other;  
  
Query Match 12.7%; Score 98; DB 24; Length 717;  
Best Local Similarity 65.6%; Pred No. 6.7e-17;  
Matches 143; Conservative 0; Mismatches 75; Indels 0; Gaps 0;  
  
QY 552 GATTGATGGCATTGACAAATCTTGTTGTCGGAAGGAGTTTCATCGTCTGGGCAGC 611  
II III III III III III III III III III III III III III III III III  
Db 28 GACCTGGTGGCATAGAGAACCTCTGCTCTCGCAGGAGTTTCATCGTCTGAGGCTGC 87  
  
QY 612 CTCAGCAAGCTCTCGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCACAGCTC 671  
II III III III III III III III III III III III III III III III III  
Db 88 CTTCAAGCTCACCAGAAGGGCTCGCAGCAGAGGATGTTTTTTCTGTTCAGATATG 147  
  
QY 672 CTGCTATACAGCGCGGGGCTGACGGCTCCCAATCAGTTTAAAGTCCACGGCAGCTC 731  
II III III III III III III III III III III III III III III III III  
Db 148 TTGCTGTACAGAGCAAGAGGTTGAGGAGCAGCCACTTCGGATCCGGGCTCCTT 207  
  
QY 732 CGCTCTATGCGATGACGATTGAGGAGCGCAAGACGA 769  
II III III III III III III III III III III III III III III III III  
Db 208 CCCCTCCAAGGATGCTGTGTGAAGAAAGTGATAACGA 245  
  
RESULT 9  
ABLO3867  
ID ABL03867 standard; cdna; 1747 BP.  
XX  
AC ABL03867;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 6083.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR P-PSDB; ABB59764.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1; SEQ ID NO 6083; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA

CC sequences (ABLI01840-ABLI16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1747 BP; 431 A; 492 C; 461 G; 363 T; 0 other;  
  
Query Match 12.6%; Score 97.2; DB 23; Length 1747;  
Best Local Similarity 50.8%; Pred No. 1.7e-16;  
Matches 259; Conservative 0; Mismatches 248; Indels 3; Gaps 1;  
  
QY 171 GATTACCAAGAATCGCGCATCTCATGTGAAGAACATTTACGGGCATGAAGCACCTGGCG 230  
II III III III III III III III III III III III III III III III III  
Db 516 GAAGCCCACCGCATCGGAGACGTCATGATGAAGACATCGCGCTCTGCCCATCTACGAT 575  
  
QY 231 GCTCACCCTGTGGAACACACAGCAGGCCCTTGGAGGCCCTTGGAGAATGGAATCAAGAGCTCC 290  
II III III III III III III III III III III III III III III III III  
Db 576 GAGTACGTGCAGACGACCTGGACATCTTGCACTGTATGAACGACATGTACGAAGCGCAT 635  
  
QY 291 CGCGCGCTGGAGAATCTCTGCAGAGACTTTGAGCTGCAGAAGGTGTGTACCTACCGCTC 350  
II III III III III III III III III III III III III III III III III  
Db 636 GAACGCTTCCTCAGGTGTACAGAGGATTTGAGCAGCAAAAGGTTTGTCTATCTACCCATC 695  
  
QY 351 AACACCTTCTCTCTCGGCCACTGCACCGGCTCATGTCATACAAGCAGTCTCTGGAGCGG 410  
II III III III III III III III III III III III III III III III III  
Db 696 GCGGAACCTTCTACTGAAGCCCTCAACGCCCTGTTGCACTACCAACTGATCTTAGAGCGG 755  
  
QY 411 CTGTGCAAAACACCACCGCCGAGCCAGCCGACTTCAGGGAGCTGCGGAGCCCTTTGGCA 470  
II III III III III III III III III III III III III III III III III  
Db 756 CTCTCGACTACTATGGGAGGAGCATATCGACTATGCGGATGCTATGGCCGCTGCACCAC 815  
  
QY 471 GAGATCAGGAGATGGTGGCACAGCTCCACAGTCCAGTATCAAGATGGAGAATTTCCAG 530  
II III III III III III III III III III III III III III III III III  
Db 816 TTGCTCGTTCGCGACCAACCAAGGTTATTAGTCCCGACTTCCCGACTCTGCAAACTTTGTG 875  
  
QY 531 AAGCTGCAGAACTCAAGAAAGATTTGATTTGCAATTTGTTGTTCCCGGAAGG 590  
II III III III III III III III III III III III III III III III III  
Db 876 GAGCTGTGTGAACT--GCAACGCGACATCAACTTCGAGCAGTTGGTTCAGCCCCCATCGC 932  
  
QY 591 GAGTTTCATCCCTCTGGGAGCCCTCAGCAAGCTCTCGGGAAGGGGCTCCAGCAGCGCATG 650  
II III III III III III III III III III III III III III III III III  
Db 933 CGCCTCATCCGCGAGGATGCCCTTCTGAAGCACTCCAAAGCGCGTCTGCAGCAGAGGATG 992  
  
QY 651 TTCTTCTCTTCAACGAGCTCTCTGCTATAC 680  
II III III III III III III III III III III III III III III III III  
Db 993 TTCTTCTTGTCTCCGACTGCTGCTCTAC 1022  
  
RESULT 10  
ABLO3866/C  
ID ABL03866 standard; cdna; 4150 BP.  
XX  
AC ABL03866;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 6080.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX



XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 11-JAN-2000; 2000JP-0118776.  
XX 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI: 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
XX full-length cDNAs defined in the specification, and for the detection  
XX and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
XX Claim 1; SEQ ID 4702; 2537pp + CD ROM; English.  
XX The present invention describes primer sets for synthesizing 5602  
XX full-length cDNAs defined in the specification. Where a primer set  
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
XX to the complementary strand of a polynucleotide which comprises one of  
XX the 5602 nucleotide sequences defined in the specification, where the  
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX of an oligonucleotide comprising a sequence complementary to the  
XX complementary strand of a polynucleotide which comprises a 5'-end  
XX sequence and an oligonucleotide comprising a sequence complementary to a  
XX polynucleotide which comprises a 3'-end sequence, where the  
XX oligonucleotide comprises at least 15 nucleotides and the combination of  
XX the 5'-end sequence/3'-end sequence is selected from those defined in  
XX the specification. The primer sets can be used in antisense therapy and  
XX in gene therapy. The primers are useful for synthesizing polynucleotides,  
XX particularly full-length cDNAs. The primers are also useful for the  
XX detection and/or diagnosis of the abnormality of the proteins encoded by  
XX the full-length cDNAs. The primers allow obtaining of the full-length  
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
XX represent oligonucleotides, all of which are used in the exemplification  
XX of the present invention.  
XX Sequence 665 BP; 209 A; 138 C; 154 G; 161 T; 3 other;  
Query Match 10.9%; Score 83.6; DB 22; Length 665;  
Best Local Similarity 51.3%; Pred. No. 6e-13;  
Matches 194; Conservative 0; Mismatches 184; Indels 0; Gaps 0;  
QY 338 TTACCTACCGCTCAACACCTTCTCTCTCGGGCCACTGACCGGCTCATCTACAGCA 397  
Db 27 TAATCGGCCCTCAAGCACTACCTGTCTCAAGCGGTTTCAGAGGATCCCGACGACGGT 86  
QY 398 GGTCTCTGAGCGGCTGTGCAACACCAACCCGCCGAGCCGACCTTCAGGACGTGCGG 457  
Db 87 GTTGCTGACAGATATTTCAGAGATCTCATAGAAGATGCTGGAGATTACAGACACTCA 146  
QY 458 AGCCGCTTTGGCAGAGATCAGGAGATGTGGCAGAGCTCCACGCTTACGATGATCAAGAT 517  
Db 147 AGATGCCCTTGCTGTTGTATAGAGTAGTCCCAACCAACGCGCCATGACACCATGAAGCAAG 206  
QY 518 GGAGAAATTCAGAGAGCTCCAGCACTCAAGAAAGATTTGATTGGCATTGACAATCTTGT 577

Db 207 AGACAACTTTTCAGAAACTTATGCAAAATTCAGTACAGCTTAAATGGACACCATGAAATTGT 266  
QY 578 GGTTCGGGAAGGAGTTCATCGTCTGGCGACGCTCAGCAAGCTCTCGGGAAGGGCT 637  
Db 267 GCAGCTGTGTCGGGTTTCTCAAGAGGAATTCGTGAAGCTCTCTCGGAAAGTGAT 326  
QY 638 CCAGCAGCGCATGTTCTCTCTTCAACGACGCTCTGTATACACGAGCGGGGGCTGAC 697  
Db 327 GCAACCTCGAATGTTTCTCTGTTTAATGATGCCCTGCTGTATACAACACCATGCGAGTC 386  
QY 698 GGCCTCCAATCAGTTAA 715  
Db 387 TGGGATGTATAAACTGAA 404  
RESULT 13  
ABK49898  
ID ABK49898 standard; cDNA; 1296 BP.  
XX  
AC ABK49898;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Human cDNA encoding Faciogenital dysplasia 1-like protein.  
XX  
KW Human; ss: gene; faciogenital dysplasia; FGDL-like protein;  
KW Aarskog syndrome; X-linked developmental disorder; cancer; obesity;  
KW guanine nucleotide exchange factor; overweight; anorexia; cachexia;  
KW wasting disorder; appetite suppression; appetite enhancement; satiety;  
KW modulation of body weight; eating disorder; bulimia; hypertension;  
KW type 2 diabetes; coronary artery disease; hyperlipidaemia;  
KW stroke; gallbladder disease; gout; osteoarthritis; sleep apnea;  
KW respiratory problem; thrombolic disease; polycystic ovarian syndrome;  
KW reduced fertility; complication of pregnancy; menstrual irregularity;  
KW hirsutism; stress incontinence; depression.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
CDS 1..1296  
FT /\*tag= a  
FT /product= "FGDL-like protein"  
FT /partial  
FT /note= "No stop codon shown"  
XX  
PN WO200224731-A2.  
XX  
PD 28-MAR-2002.  
XX  
PF 24-SEP-2001; 2001WO-EP11009.  
XX  
PR 25-SEP-2000; 2000US-234983P.  
XX  
PA (FARB ) BAYER AG.  
XX  
PI Kossida S;  
XX  
DR WPI: 2002-383181/41.  
XX P-PSDB; AAU80165.  
XX  
PT New human faciogenital dysplasia (FGDL)-like proteins, regulators of  
XX which are useful for preventing and treating cancer, obesity, cachexia,  
XX depression, diabetes, hypertension and stroke -  
XX  
PS Claim 1; Fig 1; 78pp; English.  
XX  
CC The invention relates to a purified human faciogenital dysplasia (FGDL)-  
XX like protein and its encoding cDNA. Faciogenital dysplasia (also known as  
XX Aarskog syndrome) is an X-linked developmental disorder and is a guanine  
XX nucleotide exchange factor. Also included are a hybridisable sequence,  
XX fragment, derivative, allelic variant, or a sequence which deviates  
XX from the cDNA (or sequence 50% identical to it) due to degeneration of

CC the genetic code, an expression vector comprising the nucleotide  
 CC sequence, a host cell containing the vector, modulators of the activity  
 CC of the protein, fusion proteins with the FGDI-like protein, and screening  
 CC for agents which modulate/reduce the activity of the protein.  
 CC The FGDI-like protein and cDNA are useful for screening for agents which  
 CC decrease the activity of FGDI-like protein. The cDNA is useful for  
 CC detecting a polynucleotide encoding a FGDI-like protein in a biological  
 CC sample. The protein modulators and isolated agents are useful for  
 CC treating a FGDI-like protein dysfunction related disease, such as cancer,  
 CC obesity, overweight, anorexia, cachexia, wasting disorders, appetite  
 CC suppression, appetite enhancement, increases or decreases in satiety,  
 CC modulation of body weight, and/or other eating disorders such as  
 CC bulimia, obesity/overweight-associated co-morbidities including  
 CC hypertension, type 2 diabetes, coronary artery disease, hyperlipidaemia,  
 CC stroke, gallbladder disease, gout, osteoarthritis, sleep apnea and  
 CC respiratory problems, endometrial, breast, prostate, colon cancer,  
 CC thrombotic disease, polycystic ovarian syndrome, reduced fertility,  
 CC complications of pregnancy, menstrual irregularities, hirsutism, stress  
 CC incontinence and depression. The coding sequence of FGDI-like protein  
 CC polynucleotide is useful for generating antisense oligonucleotides or  
 CC ribozymes. These antisense oligonucleotides are useful for modulating  
 CC FGDI-like protein gene expression. The FGDI-like protein is useful for  
 CC generating antibodies against FGDI-like protein amino acid sequences and  
 CC for use in various assay systems. The present sequence is the cDNA  
 CC encoding the human FGDI-like protein.  
 XX  
 SQ Sequence 1296 BP; 415 A; 273 C; 281 G; 327 T; 0 other;

Query Match 10.9%; Score 83.6; DB 24; Length 1296;  
 Best Local Similarity 51.3%; Pred. No. 8.1e-13;  
 Matches 194; Conservative 0; Mismatches 184; Indels 0; Gaps 0;  
 QY 338 TTACCTACCGCTCAACACCTTCTCTCGGCGCACTGCACCGGCTCATCTACCTACAAGCA 397  
 DB 18 TAATCTGGCCCTCAGCACTACTCTGCTCAGCGGTTTCAGAGATCCCCAGTACAGGCT 77  
 QY 398 GGTCTCGGAGCGGCTGTGCAACACACACCGCGGACCGGCTTTCAGGACTGCGC 457  
 DB 78 GTTGTGACAGATATTATTGAAGAATCTCATAGAAGATGCTGGAGATTACAGAGACACTCA 137  
 QY 458 ACCCGCTTGGCAGAGATCAGGAGATGTCGCACAGCTCCACGCTCCAGTATCAAGAT 517  
 DB 138 AGATGCCCTTGGCTTTTATAGAGATGCCAACCCAGCCCAATGACACCATGAAGCAAGG 197  
 QY 518 GGAGATTTCCAGAGCTCAGCAACTCAAGAAAGATTTGATTTGGATTTGACAACTCTTGT 577  
 DB 198 AGACAACTTTCAGAACTTATGCAATTCAGTACAGCTTAATGGACACCATGAATTTGT 257  
 QY 578 GGTTCGGGAAGGAGTTTCATCCGCTCGGCGAGCTCAGCAAGCTCTCGGGGAAGGGGCT 637  
 DB 258 GCAGCCTGCTCGGGTTTTTCTCAAAGAGGAATTTCTGATGAAGCTGCTCGGAAAGTGAT 317  
 QY 638 CCAGCAGCGCATGCTTCTCTTCAACGACGTCCTGCTATATACAGCGCGGGGCTGAC 697  
 DB 318 GCAACCTCAAAATGTTTCTCTGTTAATGATGCCCTGCTGTATACAAACACAGTCACTC 377  
 QY 698 GGCCTCCAACTAGTTAA 715  
 DB 378 TGGATGTATAACTGAA 395

RESULT 14  
 ID AAH15014  
 XX AAH15014 standard; cDNA; 1821 BP.  
 AC AAH15014;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:12971.  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX

OS Homo sapiens.  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-01161126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 12971; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 1821 BP; 594 A; 344 C; 369 G; 514 T; 0 other;

Query Match 10.9%; Score 83.6; DB 22; Length 1821;  
 Best Local Similarity 51.3%; Pred. No. 9.4e-13;  
 Matches 194; Conservative 0; Mismatches 184; Indels 0; Gaps 0;  
 QY 338 TTACCTACCGCTCAACACCTTCTCTCGGCGCACTGCACCGGCTCATCTACCTACAAGCA 397  
 DB 27 TAATCTGGCCCTCAGCACTACTCTGCTCAGCGGTTTCAGAGATCCCCAGTACAGGCT 86  
 QY 398 GGTCTCGGAGCGGCTGTGCAACACACCGCGGACCGGCTTCAGGACTGCGC 457  
 DB 87 GTTGTGACAGATATTATTGAAGAATCTCATAGAAGATGCTGGAGATTACAGAGACACTCA 146  
 QY 458 AGCCGCTTTGGCAGAGATCAGGAGATGGTGGCAGAGTCCACAGCTCCAGGTACGATGATCAAGAT 517  
 DB 147 AGATGCCCTTGTCTGTTTATAGAGTGAAGCAACCCAGCCCAATGACACCATGAAGCAAGG 206  
 QY 518 GGAGATTTCCAGAGCTCAGGAACTCAAGAAAGATTTGATTTGGCATTTGCAACTCTTGT 577  
 DB 207 AGACAACTTTCAGAAACCTTATGCAAAATTCAGTACAGCTTAAATGAGCAACCATGAAATTTGT 266

```
QY 578 GGTTCGGGAAGGAGTTTCATCCGTCTGGCAGCCCTCAGCAAGCTCTCGGGGAAGGGGT 637
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 267 GCAGCCTGCTCGGGTTTTCTCAAGAAGGAATTCGTGATGAAGCTGTCTCGGAAGTGAT 326
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 638 CCAGCAGCGCATGTTCTTCTGTTCAACGACGTCCTGCTATACACAGCGCGGGGCTGAC 697
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 327 GCAACCTCGAATGTTTTTCTGTTTAAATGATGCCCTGCTGTATACAACACCAAGTCAGTC 386
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 698 GGCCTCCAATCAGTTTAA 715
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 387 TGGGATGTATAACTGAA 404
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
ABN59933
ID ABN59933 standard; cDNA; 3610 BP.
XX
AC ABN59933;
XX
DT 28-JUN-2002 (first entry)
XX
DE Novel human coding sequence SEQ ID NO: 344.
XX
KW Human; antianemic; vulnery; antiinflammatory; immunomodulator;
KW antifertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200222660-A2.
XX
PN 21-MAR-2002.
XX
PD 10-SEP-2001; 2001WO-US26015.
XX
PF 11-SEP-2000; 2000US-0659671.
XX
PR (HYSE-) HYSEQ INC.
XX
PA Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
PI
XX WPI; 2002-292408/33.
DR P-PSDB; ABB97520.
XX
XX An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis -
PT
XX Claim 1; SEQ ID NO 344; 509pp; English.
XX
XX The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a coding sequence of the
CC invention.
XX
SQ Sequence 3610 BP; 1205 A; 674 C; 745 G; 986 T; 0 other;

Query Match 10.9%; Score 83.6; DB 24; Length 3610;
Best Local Similarity 51.3%; Pred. No. 1.3e-12;
Matches 194; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 338 TTACCTACCGCTCAACACCTTCTCTCGGGCCACTGCACCGGCTCATGCACCTACAGCA 397
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1271 TAATCTGGCCCTCAAGCACTACTCTCTCAAGCGGGTTCAGAGGATCCCCCAGTACAGGCT 1330
```

```
QY 398 GTTCCTGGAGCGGCTGTGCAAAACACACCCGCCGACCCGACTTTCAGGGACTGCGG 457
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1331 GTTGTGACAGATTATTTGAAGAAATCTCATAGAAGATGCTGGAGATTACAGAGACACTCA 1390
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 458 AGCCGCTTTGGCAGAGATCACGGAGATGGTGGCACAGCTCCACGGGTACGATCAAGAT 517
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1391 AGATGCCCTTGTGTTGTTTATAGAGGTAGCCAAACCATGCCAATGCACCATGAAGCAAGG 1450
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 518 GGAGAAATTTCCAGAAGCTGCACGAACCTCAAGAAAGATTTTGATTGGCATTGACAATCTTCT 577
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1451 AGACAACCTTTCAGAAACTTATGCAAAATTCAGTACAGCTTAATGGACACCATGAATTTGT 1510
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 578 GGTTCGGGAAGGAGTTTCATCCGTCTGGGCAGCCCTCAGCAAGCTCTCGGGGAAGGGGT 637
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1511 GCAGCCTGCTCGGGTTTTTCTCAAGAAGGAATTTCTGATGAAGCTGTCTCGGAAAGTGAT 1570
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 638 CCAGCAGCGCATGTTCTTCTGTTCAACGACGTCCTGCTATACACAGCGCGGGGCTGAC 697
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1571 GCAACCTCGAATGTTTTTCTGTTTAAATGATGCCCTGCTGTATACAACACCAAGTCAGTC 1630
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 698 GGCCTCCAATCAGTTTAA 715
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1631 TGGGATGTATAACTGAA 1648
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: December 12, 2002, 05:06:33
Job time : 326 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2002, 04:45:14 ; Search time 72 Seconds  
(without alignments)  
3275.478 Million cell updates/sec

Title: US-09-555-342b-1\_COPY\_1733\_2501

Perfect score: 769  
Sequence: 1 tcaactcgtggttcagagc.....attgaggagagcgagacga 769

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_NA:\*

1: /cgn2.6/ptodata/2/ina/5A\_COMB.seq:\*

2: /cgn2.6/ptodata/2/ina/5B\_COMB.seq:\*

3: /cgn2.6/ptodata/2/ina/6A\_COMB.seq:\*

4: /cgn2.6/ptodata/2/ina/6B\_COMB.seq:\*

5: /cgn2.6/ptodata/2/ina/PCUS\_COMB.seq:\*

6: /cgn2.6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40.8	5.3	4776	US-08-852-401-1	Sequence 1, Appli
2	39.6	5.1	4722	US-08-979-608A-14	Sequence 14, Appli
3	37.2	4.8	44377	US-08-804-227C-7	Sequence 7, Appli
4	37.2	4.8	44377	US-08-804-198-1	Sequence 1, Appli
5	36.4	4.7	2824	US-09-010-928B-3	Sequence 3, Appli
6	36.4	4.7	2214	US-08-864-038A-1	Sequence 1, Appli
7	36	4.7	3331	US-08-864-038A-2	Sequence 2, Appli
8	36	4.7	3331	US-08-864-038A-4	Sequence 4, Appli
9	35.4	4.6	4403765	US-09-552-322-1	Sequence 1, Appli
10	35.4	4.6	4411529	US-09-103-840A-2	Sequence 2, Appli
11	35.4	4.6	4411529	US-09-103-840A-1	Sequence 1, Appli
12	34.6	4.5	657	US-09-527-345-3	Sequence 3, Appli
13	34.6	4.5	1029	US-08-911-853-1	Sequence 1, Appli
14	34.6	4.5	1029	US-09-479-409-1	Sequence 1, Appli
15	34.6	4.5	1029	US-09-479-453-1	Sequence 1, Appli
16	34.6	4.5	4377	US-08-911-853-28	Sequence 28, Appli
17	34.6	4.5	4377	US-09-479-409-28	Sequence 28, Appli
18	34.6	4.5	4377	US-09-479-453-28	Sequence 28, Appli
19	34.6	4.5	10348	US-08-457-273B-41	Sequence 41, Appli
20	34.6	4.5	10348	US-08-556-419-13	Sequence 13, Appli
21	34.6	4.5	10348	US-09-041-886-14	Sequence 14, Appli
22	34.6	4.5	10366	US-08-246-982A-5	Sequence 5, Appli
23	34.6	4.5	10366	US-08-453-265-5	Sequence 5, Appli
24	34.4	4.5	7042	US-09-092-508-1	Sequence 1, Appli
25	34.4	4.5	7042	US-09-435-115-1	Sequence 1, Appli
26	34.4	4.5	7042	US-09-098-310-1	Sequence 1, Appli
27	34.4	4.5	7042	US-09-690-364-21	Sequence 21, Appli

c 28	34.4	4.5	7075	4	US-09-092-508-15	Sequence 15, Appli
c 29	34.4	4.5	7075	4	US-09-435-115-15	Sequence 15, Appli
c 30	34.2	4.4	1331	4	US-09-335-983-14	Sequence 14, Appli
31	34.2	4.4	4195	1	US-08-340-011-1	Sequence 1, Appli
32	34.2	4.4	4195	3	US-08-901-710-1	Sequence 1, Appli
33	34.2	4.4	4416	3	US-08-795-430-1	Sequence 1, Appli
34	34.2	4.4	4416	4	US-09-355-700-1	Sequence 1, Appli
35	34.2	4.4	4416	4	US-08-601-132-36	Sequence 36, Appli
36	34.2	4.4	4425	1	US-08-222-616-31	Sequence 31, Appli
37	34.2	4.4	4425	4	US-08-446-648-31	Sequence 31, Appli
38	34.2	4.4	4425	5	PCT-US95-04228-31	Sequence 31, Appli
39	34.2	4.4	4795	1	US-08-340-011-3	Sequence 3, Appli
40	34.2	4.4	4795	3	US-08-901-710-3	Sequence 3, Appli
c 41	34.2	4.4	5349	4	US-09-068-101-7	Sequence 7, Appli
42	34.2	4.4	6548	3	US-08-894-440-1	Sequence 1, Appli
43	34.2	4.4	6548	3	US-08-817-188-2	Sequence 2, Appli
44	34.2	4.4	6548	4	US-09-458-093-1	Sequence 1, Appli
45	34.2	4.4	9108	4	US-08-446-648-45	Sequence 45, Appli

## ALIGNMENTS

RESULT 1  
US-08-852-401-1  
; Sequence 1, Application US/08852401  
; Patent No. 5976836  
; GENERAL INFORMATION:  
; APPLICANT: Weber, J. Mark  
; APPLICANT: Hessler, Paul E.  
; APPLICANT: Larsen, Peter E.  
; APPLICANT: Luu, B. Minh  
; TITLE OF INVENTION: Methods and Compositions for Enhancing  
; TITLE OF INVENTION: Erythromycin Production  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rockey, Milnamow & Katz, Ltd.  
; STREET: 2 Prudential Plaza, 180 N. Stetson, Suite  
; STREET: 4700  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: U.S.A.  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/852,401  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mueller, Lisa L.  
; REGISTRATION NUMBER: 38,978  
; REFERENCE/DOCKET NUMBER: FER2159P00300S  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-616-5400  
; TELEFAX: 312-616-5460  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4776 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-852-401-1

Query Match 5.3%; Score 40.8; DB 2; Length 4776;  
Best Local Similarity 43.9%; Pred. No. 0.058;  
Matches 174; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

QY 280 TCAAGAGCTCCCGCGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAGGTTGT 339

Db 797 TCAACCGCGGGCGGCATCAAGAGGACACCGCCAGCGCGTGTGGCGCGTCAACG 856  
QY 340 ACCTACCGCTCAACACCTTCTCCCTCGCGCCACTGCACCGGCTCATGCACTACAAGCAGG 399  
Db 857 AGATCGCGCACACCGCGCGCATAGGTGCGCTGCGCGCACCGGGGTAATCGGCGCTGC 916  
QY 400 TCTGAGCGGGTGTGCAACACACCGCGCGAGCCCGGACCTTCAGGGACTGCGCGAG 459  
Db 917 TGGTCCGAGCTGTCCACACCGGTTTCCCGGCTTCCCGAGGCGCTGGAGGCGCGG 976  
QY 460 CGCTTTGCGAGAGTACGAGAGTGTGGACAGCTGCACCGTACGATGATCAAGATGG 519  
Db 977 CGTCCGCGCGGCTACGCTCGCTGTGCAACACCGCGCTGCGGATGACGAGGAGG 1036  
QY 520 AGNATTTCCAGAGCTGCACGACTCAAGAAAGATTTGATTTGGCATTCATCTTGGG 579  
Db 1037 ACTAGCTCCGGATGTTCATCGCGCGCGGCTGTGAGGCAATGCTTCTGCTGCGCGGAGA 1096  
QY 580 TTCCGGGAAGGAGTTTCATCCGCTGTGGGCGAGCTCAGCAAGCTCTCGGGGAAGGGCTCC 639  
Db 1097 TCGCCAACACCGAGGCGGAGCGGATCAGCGGAGCTACTACGAGAAGCTGCTGGCGG 1156  
QY 640 AGCAGCGCATGTTCTTCCTGTTTCAACGAGCGTCTCTCG 675  
Db 1157 ACGGCGTGGCGCATGTTCTTCGTCAACGCGCGCGC 1192

RESULT 2

US-08-979-608A-14  
; Sequence 14, Application US/08979608A  
; Patent No. 6355451  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; Lees, Robert S.  
; Law, Simon W.  
; Arjona, Anibal A.  
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN  
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND  
; TREATING ATHEROSCLEROSIS  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/979,608A  
; FILING DATE: 26-NOV-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/048,547  
; FILING DATE: 03-JUN-1997  
; APPLICATION NUMBER: US 60/031,930  
; FILING DATE: 27-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Louis  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59810)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4722 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single

; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 61...1731  
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-08-979-608A-14  
Query Match 5.1%; Score 39.6; DB 4; Length 4722;  
Best Local Similarity 47.5%; Pred. No. 0.13;  
Matches 154; Conservative 0; Mismatches 164; Indels 6; Gaps 1;  
QY 218 GAAGACACCTGGCGGCTCACCTGTGGAAGCACACGAGGCGCTTTGGAGGCCCTTGGAGAATGG 277  
Db 618 GGAGAAGCTGGCGGCTGTGTGCAAGAATATGCGGAACCTGCTCGAGGAGACCGCAACTC 677  
QY 278 ATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAAGAGACTTTGAGCTGCAGAAAGTGTG 337  
Db 678 GCAGAAGCAGATGAAGCTGCTGCAAGAAGCAGAGCCAGCTGGTGCAGGAGAGG---- 733  
QY 338 TTACTACCGCTCAACACCTTCTCTCTCGGGCCACTGCACCGGCTCATGCACCTACAAGCA 397  
Db 734 --ACACCTGCTGGCGGAGCACAGCAAGCCATCTGCGCCGACAGCAAGCTCGAGAGCT 791  
QY 398 GGTCTCTGGAGCGGCTGTGCAAAACACACCGCGGAGCCGCGGACTTTCAGGGACTGCGG 457  
Db 792 GTGCGCGGAGCTGACGCGGCACACCGCTCGCTCAAGGAAGAAGCTGTGCACGAGCGCG 851  
QY 458 AGCGCTTTGGCAGAGATCAGGGAGATGTTGCCACAGCTCCACGGTAGCATGATCAAGAT 517  
Db 852 AGAGGAGGAGAGAGCGCAAGGAGGTGACGTCAACACTTCCAGATGACGCTCAACGACAT 911  
QY 518 GGAGAATTTCCAGAAGCTGCACGA 541  
Db 912 TCAGCTGCAGATGGAGCAGCACAA 935

RESULT 3

US-08-804-227C-7  
; Sequence 7, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas, G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4437 base pairs  
; TYPE: nucleic acid



```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-227C-7

Query Match
Best Local Similarity 4.8%; Score 37.2; DB 2; Length 44377;
Matches 162; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

QY 147 GGCCGCTCAAAATGCCCAATCAGAGATTACCAAGAAATCGCGGATGTCATGCTGAAGAAC 206
Db 29515 GACTTCGCGACCGCTTACCGGAGCGGACGACGCTCCCGCGGACGCTCGTCTGAAC 29574

QY 207 ATTCAGGCGATGAAGACCTTCGCGGCTACCTGTGGAACACAGCAGCGCTTGGAGGCC 266
Db 29575 TCGCTCACCAGGAGTTCGTGGACGCTCCCTCGGCTGCTCCGCGCGCGGTTTC 29634

QY 267 CTGGAGATGGAATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAGAGACTTTGAGCTG 326
Db 29635 CTGGAGCTGGCAAGACCGACGCTCCGGGACCCCGGAGCGGATCGCCGCCCAACACCCCGG 29694

QY 327 CAGAAGGTGTGTACCTACCGCTCAACACCTTCCTCCTCGCGGCTGACACCGGCTCATG 386
Db 29695 GTGCGCTACCGGCGTTCACCTCAACGAGCGGCGGACCGGCTCGCGCGGCTGCTG 29754

QY 387 CACTACAAGCAGGTCTGTGAGCGGTGTGCAACACCAACCGCGGAGCGACGCGGCTTC 446
Db 29755 CGGGAACATGATGAGCTGTTCGCGCGCGGCTGCTGCAACCGCTGCGCGCTCAACCCAC 29814

QY 447 AGGAGCTCGCGGCGGCTTGGCAGAGATCACGGAGATGGTGCAAGCTTCCACGGTACG 506
Db 29815 GACGTGCGCGCGGCGGCGGACGCTCCGCGCACCATCAGCCAGCGCGCGGCAACCGGAAG 29874

QY 507 ATGATCAAGA 516
Db 29875 CTCGTCCTGA 29884
```

## RESULT 4

```
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuntoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostock, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN USA
; COUNTRY: IN
; ZIP: 46285
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-198-1

Query Match
Best Local Similarity 4.8%; Score 37.2; DB 2; Length 44377;
Matches 162; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

QY 147 GGCCGCTCAAAATGCCCAATCAGAGATTACCAAGAAATCGCGGATGTCATGCTGAAGAAC 206
Db 29515 GACTTCGCGACCGCTTACCGGAGCGGACGACGCTCCCGCGGACGCTCGTCTGAAC 29574

QY 207 ATTCAGGCGATGAAGACCTTCGCGGCTACCTGTGGAACACAGCAGCGCTTGGAGGCC 266
Db 29575 TCGCTCACCAGGAGTTCGTGGACGCTCCCTCGGCTGCTCCGCGCGCGGTTTC 29634

QY 267 CTGGAGATGGAATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAGAGACTTTGAGCTG 326
Db 29635 CTGGAGCTGGCAAGACCGACGCTCCGGGACCCCGGAGCGGATCGCCGCCCAACACCCCGG 29694

QY 327 CAGAAGGTGTGTACCTACCGCTCAACACCTTCCTCCTCGCGGCTGACACCGGCTCATG 386
Db 29695 GTGCGCTACCGGCGTTCACCTCAACGAGCGGCGGACCGGCTCGCGCGGCTGCTG 29754

QY 387 CACTACAAGCAGGTCTGTGAGCGGTGTGCAACACCAACCGCGGAGCGACGCGGCTTC 446
Db 29755 CGGGAACATGATGAGCTGTTCGCGCGCGGCTGCTGCAACCGCTGCGCGCTCAACCCAC 29814

QY 447 AGGAGCTCGCGGCGGCTTGGCAGAGATCACGGAGATGGTGCAAGCTTCCACGGTACG 506
Db 29815 GACGTGCGCGCGGCGGCGGACGCTCCGCGCACCATCAGCCAGCGCGCGGCAACCGGAAG 29874

QY 507 ATGATCAAGA 516
Db 29875 CTCGTCCTGA 29884
```

## RESULT 5

US-09-010-928B-3/C

RESULT 6  
US-08-864-038A-1/c  
; Sequence 1, Application US/08864038A  
; Patent No. 6001592  
; GENERAL INFORMATION:  
; APPLICANT: Kunio NAKASHIMA et al.

RESULT 7  
US-08-864-038A-2/c  
; Sequence 2, Application US/08864038A  
; Patent No. 6001592  
; GENERAL INFORMATION:  
; APPLICANT: Kunio NAKASHIMA et al.  
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR  
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID  
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING  
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD  
; TITLE OF INVENTION: TO SAID POLYPEPTIDE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: 812-5 Hirano



```
; ORGANISM: Homo sapiens
US-09-552-322-1

Query Match          4.6%; Score 35.4; DB 4; Length 2132;
Best Local Similarity 55.2%; Pred. No. 1.4;
Matches 69; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 607 GCAGGCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTTCAACG 666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 176 GGGGGCGGCTGAGCTCTCGGCTCGGCTTCGACGAGCGGCAACTTCTCGTGTCTATCC 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 667 AGCTCTGCTATACAGCAGCGCGGGGCTGACGGCCTCCAATCAGTTTAAAGTCCACGGGC 726
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 236 GCGCGGTGGAGGAGAGCAGGCGGGGCTGTACACCTGCAACCTGCACCATCACTACTGCC 295
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 727 ACCTC 731
    ||| |||
DB 296 ACCTC 300

RESULT 10
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          4.6%; Score 35.4; DB 4; Length 4403765;
Best Local Similarity 52.3%; Pred. No. 43;
Matches 78; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 345 CCGCTCAACACCTTCTCTCGGGCACTGCACCGGCTCATGCAAGCAGGTCTCTG 404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 391026 CTGCGGTAGACGTTATCCGGGTATGCGAGTAGTTGACATGATCGACACCCAGGACCTG 391085
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 405 GAGCGGCTGTGCAACACACCCCGCGGACCTTCAGGACTTCAGGAGCTGCCGAGCCGCT 464
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 391086 TCCCGCAGGCGGAGTCTCTCCCGGCGACCTCACCGAGGTTTCGGTCCGCGCCGAG 391145
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 465 TTGGCAGAGATCAGGAGATGTGGCACA 493
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 391146 TGTGAGGCTCGCGGTGGGTGAAGCACA 391174
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
```

```
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match          4.6%; Score 35.4; DB 4; Length 4411529;
Best Local Similarity 52.3%; Pred. No. 43;
Matches 78; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 345 CCGCTCAACACCTTCTCTCGGGCACTGCACCGGCTCATGCAAGCAGGTCTCTG 404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 390971 CTGCGGTAGACGTTATCCGGGTATGCGAGTAGTTGACATGATCGACACCCAGGACCTG 391030
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 405 GAGCGGCTGTGCAACACACCCCGCGGACCTTCAGGACTTCAGGAGCTGCCGAGCCGCT 464
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 391031 TCCCGCAGGCGGAGTCTCTCCCGGCGACCTCACCGAGGTTTCGGTCCGTCGCGCCGAG 391090
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 465 TTGGCAGAGATCAGGAGATGTGGCACA 493
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 391091 TGTGAGGCTCGCGGTGGGTGAAGCACA 391119
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-09-527-345-3
; Sequence 3, Application US/09527345
; Patent No. 6331413
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/527,345
; CURRENT FILING DATE: 1999-03-17
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 657
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide sequence for zsig63
; NAME/KEY: misc_feature
; LOCATION: (1)...(657)
; OTHER INFORMATION: n = A,T,C or G
US-09-527-345-3

Query Match          4.5%; Score 34.6; DB 4; Length 657;
Best Local Similarity 36.5%; Pred. No. 1.3;
Matches 61; Conservative 9; Mismatches 97; Indels 0; Gaps 0;

QY 218 GAAGCACCTGCGGGCTCACCTGTGGAAGCACAGCGAGCGCTTGGAGCGCTGGAGATGG 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 439 GARCNGCNGCNGCNGCNGCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGTNGNGCN 498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 278 AATCAAGAGCTCCCGCGGCTGGAGAACTTCTGACAGACTTTTGAGTCGACAGAGGTGTG 337
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 499 GARCNGCNGCNGARCCNGCNGTNGCNGCNGARCCNGCNGCNGCNGCNGCNGTNGNGNTN 558
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 338 TTACCTACCGCTCAACACCTTCTCTCGCGCCACTGCACCGGCTCA 384
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 559 GARCNGCNGCNGARCCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNA 605
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```





Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
c	1	762	99.1	3094	10	US-09-925-297-220	Sequence 220, App
	2	248.6	32.3	408	10	US-09-783-590-3620	Sequence 3620, Ap
	3	237.4	30.9	500	10	US-09-783-590-3621	Sequence 3621, Ap
	4	219.8	28.6	379	10	US-09-960-352-154	Sequence 154, App
	5	214.8	27.9	400	10	US-09-960-352-156	Sequence 156, App
	6	190.8	24.8	335	10	US-09-783-590-3600	Sequence 3600, Ap
	7	168.8	22.0	356	10	US-09-783-590-3575	Sequence 3575, Ap
	8	133.2	17.3	288	10	US-09-783-590-3542	Sequence 3542, Ap
	9	120.8	15.7	426	10	US-09-815-343-1490	Sequence 1490, Ap
	10	106	13.8	1718	9	US-09-764-868-51	Sequence 51, Appl
c	11	105.2	13.7	716	9	US-09-764-868-475	Sequence 475, App
	12	64.4	8.4	2686	9	US-09-764-868-48	Sequence 48, Appl
	13	64.4	8.4	3380	10	US-09-799-799-1	Sequence 1, Appl
	14	64	8.3	1091	9	US-09-764-868-316	Sequence 316, App
	15	43.4	5.6	3786	10	US-09-815-242-7865	Sequence 7865, Ap
	16	42.4	5.5	1730	9	US-09-860-670-78	Sequence 78, Appl
	17	42.4	5.5	1730	9	US-09-764-868-52	Sequence 52, Appl
	18	42.4	5.5	1730	9	US-09-764-868-476	Sequence 476, App
	19	40.8	5.3	2172	10	US-09-815-242-4038	Sequence 4038, Ap

```
|||||
Db 182 AGCACTTGCCCTGTGGAGGCGCTCAAAATGCCAAAATCAGAGATTACCAAGAAATCGG 241
QY 188 CGATGTCATGCTGAGAACATTCAGGGCATGAAGCACCTGGCGGCTCACCCTGTGGAAACA 247
Db 242 CGATGTCATGCTGAGAACATTCAGGGCATGAAGCACCTGGCGGCTCACCCTGTGGAAACA 301
QY 248 CAGCAGGCGCTTGGAGGCGCTTGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTT 307
Db 302 CAGCAGGCGCTTGGAGGCGCTTGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTT 361
QY 308 CTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACACCTTCTCTCTCGG 367
Db 362 CTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACACCTTCTCTCTCGG 421
QY 368 GCACCTGCACCGGCTCATGCACTACAAGCAGTCTCGAGCGGCTGTGCAACACCAACCC 427
Db 422 GCACCTGCACCGGCTCATGCACTACAAGCAGTCTCGAGCGGCTGTGCAACACCAACCC 481
QY 428 GCCGAGCCACGCGGCTTCAGGAGCTGCGGAGCTGCTTGGCAGAGATCACGGAGATGTT 487
Db 482 GCCGAGCCACGCGGCTTCAGGAGCTGCGGAGCTGCTTGGCAGAGATCACGGAGATGTT 541
QY 488 GGCACAGCTCCACGTCATGATCAAGATGGAGAAATTCAGAGCTGCACGAACTCAA 547
Db 542 GGCACAGCTCCACGTCATGATCAAGATGGAGAAATTCAGAGCTGCACGAACTCAA 601
QY 548 GAAAGATTGATGTCATGATGATCAATCTTGTGTTCCGGAAGGAGTTCATCCGCTCTGG 607
Db 602 GAAAGATTGATGTCATGATGATCAATCTTGTGTTCCGGAAGGAGTTCATCCGCTCTGG 661
QY 608 CAGCCTCAGCAAGCTCTCGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGA 667
Db 662 CAGCCTCAGCAAGCTCTCGGGAAGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGA 721
QY 668 CTTCTGCTATACAGCGCGGGGCTGACGGCTCCCAATCAGTTTAAAGTCCAGGGCA 727
Db 722 CTTCTGCTATACAGCGCGGGGCTGACGGCTCCCAATCAGTTTAAAGTCCAGGGCA 781
QY 728 GCTCCCGCTCTATGTCATGATGAGGAGGAGGAGCGAAGACGA 769
Db 782 GCTCCCGCTCTATGTCATGATGAGGAGGAGGAGCGAAGACGA 823
```

```
RESULT 2
US-09-783-590-3620
; Sequence 3620, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3620
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (73)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
```

```
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (340)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (352)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (356)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (399)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (402)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (406)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-3620

Query Match 32.3%; Score 248.6; DB 10; Length 408;
Best Local Similarity 85.0%; Pred. No. 2.2e-64;
Matches 347; Conservative 0; Mismatches 46; Indels 15; Gaps 6;

QY 39 GCCATGCCGGAAGCACTCAAAAGTCTCATATTCGCCGAATTTTGAACCTTTTGCACAAATTT 98
Db 1 GGCAGAGCGGAAGCACTGAAAAGTCTCATATTCGCCGAATTTTGAACCTTTTGCACAAATTT 60
QY 99 CATACTAATTTTCTCAAGGAAATTGAGCAAGCACTTGCCCTGTGGGAAGCGCGCTCAAAAT 158
Db 61 CATACTAATTTTNTCAAGAAATTTGAGCAAGCACTTGCCCTGTGGGAAGCGCGCTCAAAAT 120
QY 159 GCCAAATFCAGAGATTTACCAAGAAATCGGCGATGTCATGCTGGAAGAACATTCAGGGCATG 218
Db 121 GCCAAATFCAGAGATTTACCAAGAAATCGGCGATGTCATGCTGGAAGAACATTCAGGGCATG 180
QY 219 AAGCACCTGGCGGCTCACCTGTGGAAGCACAGCGA-GGCCTTGGAGGCCCTG---GAGAA 274
Db 181 AAGCACCTGGCGGCTCACCTGTGGAAGCACAGCGAGAGGCCCTTGGAGGCCCTTGGAGGAAAT 240
QY 275 TGAATCAAGAGCTCCCGCGGCTGGAGA----ACTTCTGCAGAGACTTTGAGCTGCAGA 330
Db 241 GGAATCAAGAGCTCCCGCGGCTGGAGAACTTTCTGCAGAGACTTTGAGCTGCAGA 300
QY 331 AGTGTGTTACCTACCGCTCAACACCTTCTCTCTGCG-----GGCAGCTGCACCGGCTCAT 385
Db 301 AGTGTGTTTACCTAACCGNTTCAACAACTTTCTCTCTGNGGGCCACTGCANCNGTCAAT 360
QY 386 -GCACCTACAAGCAGTCTCTGGAGC-GGCTGTGCAAAACACCAACCCGCCG 431
Db 361 GGCATTACAAGCAAGTTCTGGAGCGGTTGTGCAAAACANCAACCCGNCG 408
```

```
RESULT 3
US-09-783-590-3621
; Sequence 3621, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
```



NUMBER OF SEQ ID NOS: 12485  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3621  
LENGTH: 500

TYPE: DNA  
ORGANISM: Homo sapiens  
NAME/KEY: misc feature  
LOCATION: (73)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (146)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (240)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (246)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (275)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (336)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (377)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (380)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (388)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (390)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (413)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (436)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (450)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (455)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (459)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (461)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (475)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (477)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (487)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (499)  
OTHER INFORMATION: n equals a,t,g, or c

US-09-783-590-3621

Query Match 30.9%; Score 237.4; DB 10; Length 500;  
Best Local Similarity 83.3%; Pred. No. 5.3e-61;  
Matches 415; Conservative 0; Mismatches 60; Indels 23; Gaps 13;

Qy 39 GCCATGCCGGAAGCACTGAAAAGTCTCATATTCGCGAATTTTGACACTTTGCACAAATTT 98  
Db 1 GGCAGAGCGGAAGCACTGAAAAGTCTCATATTCGCGAATTTTGACACTTTGCACAAATTT 60  
Qy 99 CATACTAATTTTCTCAAGGAATTTGAGCAACGACTTGGCCCTGT-GGGAGGCCGCTCAA 157  
Db 61 CATACTAATTTTNTCAAGGAATTTGAGCAACGACTTGGCCCTGTGGGGAGGCCGCTCAA 120  
Qy 158 TGCCCAATCAGAGATTACCAAGAATCGCGATGCTCATGCTGAA-GAACATTCAGGCA 216  
Db 121 TGCCCAATCAGAGATTACCAAGAATCGCGATGCTCATGCTGAAGGAACATTCAGGCA 180  
Qy 217 TGAAGCACTTGGCGG-CTACCTGTG-GGAAGCACAGCG-AGGCCCTTGGAGGCCCTGGAGA 273  
Db 181 TGAAGCACTTGGCGGTCTCACTCTGTGGAAGCACAGCGAAGCCCTTGGAGGCCCTGGAGN 240  
Qy 274 A--TGGAAATCAAGAGCTCCCGCGG--CTGGAGAACTTCTGCAGAGACTTT-GAGCTGCA 328  
Db 241 AATGNAATTCAGAGCTCCCGCGGCTGGAGGAGNCCTTCTGCAGAGACTTTGGAGCTGCA 300  
Qy 329 GAAGGTGTG-TTACCTACCGCTCAACACTTT--CCTCCTCGCGCCACTGCACCGGCTCAT 385  
Db 301 GAAGGTGTGTTTACCTACCGCTCAACACTTTCTCTNCTGCGGGCCACTGCACCGGTTTCA 360  
Qy 386 GCACTACAAG--CAGGTCTCTGGAGCGGCTGTGCAAAACACACCCCGCGAGCCCGCA- 442  
Db 361 GCACTACAAGCGAGTCTNCTNGAGCGGNTTTCGCAAAACACACCCCGATCGNGSCCAAG 420  
Qy 443 -----CTTCAGGAGCTCCCGCGGCTTTGGCAGAGATCAGGAGATGGTGGCACACT 496  
Db 421 GCCGATTTTCAGGAGTCCCGCGG-TTTNGCAGNGTTNANGAGATTTGGGAGANNTT 479  
Qy 497 CCACGGTACGATGATCAA 514  
Db 480 CCAGGTNGTGTGTTCAA 497

## RESULT 4

US-09-960-352-154

; Sequence 154; Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Bvatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960.352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 154  
; LENGTH: 379  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 01-LIB34-020-Q1-E1-A9  
US-09-960-352-154

Query Match 28.6%; Score 219.8; DB 10; Length 379;  
Best Local Similarity 90.8%; Pred. No. 7.8e-56;  
Matches 246; Conservative 0; Mismatches 22; Indels 3; Gaps 1;

Qy 1 TCACCTTCGTGGTTTCAGAGCACAGTGGACCAAGAGGAGCCCATGCCGGAAGCACTGAAAA 60  
Db 109 TCACCTTCGTGGTTTCAGAGCACAGTGGACCAAGAGGAGTCCATGCCCGAAACCTTGAAGA 168  
Qy 61 GTCATATTCGCGAATTTTGAACCTTTGCACAAATTTTCACTAATTTTCTCAAGGAAA 120  
Db 169 GTCATATTCGCGAATTTTGAACCTTTGCACAAATTTTCACTAATTTTCTCAAGGAAA 228  
Qy 121 TTGAGCAACGACTTGGCCCTGTGGGAAGGCCGCTCAATGCCCAATCA---GAGATTACC 177  
|||||

Db 229 TTGACCAACGACTTGCCTGTGGGAAGGCCGCTCGAATGCCCATCAGAGGATTACC 288  
 Qy 178 AAGAATCGGGATGTCTATCTGAAGAACATTCAGGGCATGAAGCAGCTGGCGCTCACC 237  
 Db 289 AGAATCGGAGATGTAATGCTGAAGAGATTTCAGGGATGAAGCAACTGGCCGCCCACT 348  
 Qy 238 TGTGGAAGCACAGCGAGGCCCTTGAGAGCCCT 268  
 Db 349 TGTGGAAGCACAGCGAGGCCGCTGGAGGCGCT 379

RESULT 5  
 US-09-960-352-156  
 ; Sequence 156, Application US/09960352  
 ; Patent No. US20020137139A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Warren, Wesley C.  
 ; APPLICANT: Tao, Nengbing  
 ; APPLICANT: Byatt, John C.  
 ; APPLICANT: Mathialagan, Nagappan  
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
 ; FILE REFERENCE: 16511.006/37-21(10298)C  
 ; CURRENT APPLICATION NUMBER: US/09/960,352  
 ; CURRENT FILING DATE: 2001-09-24  
 ; NUMBER OF SEQ ID NOS: 15112  
 ; SEQ ID NO 156  
 ; LENGTH: 400  
 ; TYPE: DNA  
 ; ORGANISM: Bos taurus  
 ; OTHER INFORMATION: Clone ID: 01-LTB34-020-Q1-E2-A9  
 US-09-960-352-156

Query Match 27.98; Score 214.8; DB 10; Length 400;  
 Best Local Similarity 87.88; Pred. No. 2.5e-54;  
 Matches 258; Conservative 0; Mismatches 32; Indels 4; Gaps 2;

Qy 1 TCACCTTCGTGTTTCAGAGCACAGTGGCAAGAGGAGCGCCATGCCGGAAGCACTGAAAA 60  
 Db 107 TCACCTTCGTGTTTCAGAGCACAGTGGCAAGAGGAGCTCCATGCCGGAACCTTGAGAA 166  
 Qy 61 GTCATATATCCCGAATTTGAACTTTGCACAAATTTCACTAATTTCTCAAGGAAA 120  
 Db 167 GTCATATATCCCGAATTTGAACTTTGCACAAATTTCACTAATTTCTCAAGGACA 226  
 Qy 121 TTGAGCACGACTTGCCTGTGGGAAGGCCGCTCAAAATGCCCAATCA---GAGATTACC 177  
 Db 227 TTGAGCACGACTTGCCTGTGGGAAGGCCGCTCGAATGCCCATCAGAGGAGATTACC 286  
 Qy 178 AAGAATCGGGATGTCTATCTGAAGAACATTCAGGGCATGAAGCAGCTG-GCGGCTCAC 236  
 Db 287 ATAGAATCGGAGATGTAGTCTGAAGAACATTCAGGGATGAAGCAACTGAGCTGCTCAC 346  
 Qy 237 CTGTGGAAGCACAGCGAGGCCCTTGAGGCCCTGGAGAAATGAATCAAGAGCTCC 290  
 Db 347 TTGTGGAAGCACAGCTAGGCCCTTGAGGCCCTGGAGCTGGACATCGGCATCCTGGGCCCC 400

RESULT 6  
 US-09-783-590-3600  
 ; Sequence 3600, Application US/09783590  
 ; Patent No. US20020110850A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dillon, Patrick J.  
 ; APPLICANT: Haseltine, William A.  
 ; APPLICANT: Li, Haodong  
 ; APPLICANT: Rosen, Craig A.  
 ; APPLICANT: Ruben, Steven M.  
 ; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
 ; FILE REFERENCE: PO-16.2C1  
 ; CURRENT APPLICATION NUMBER: US/09/783,590  
 ; CURRENT FILING DATE: 2000-02-15  
 ; PRIOR APPLICATION NUMBER: 08/420,856

; PRIOR FILING DATE: 1995-04-12  
 ; PRIOR APPLICATION NUMBER: 08/346,731  
 ; PRIOR FILING DATE: 1994-11-21  
 ; NUMBER OF SEQ ID NOS: 12485  
 ; SOFTWARE: PatentIn ver. 2.0  
 ; SEQ ID NO 3600  
 ; LENGTH: 335  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (5)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (29)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (36)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (42)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (72)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (73)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (144)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (157)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (184)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (246)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (254)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (271)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (278)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (298)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (329)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (333)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; US-09-783-590-3600

Query Match 24.8%; Score 190.8; DB 10; Length 335;  
 Best Local Similarity 84.0%; Pred. No. 3.1e-47;  
 Matches 278; Conservative 0; Mismatches 42; Indels 11; Gaps 6;

Qy 39 GCCATGCCGAAGCACTGAAAGTCTCATATTCGCCGAATTTGAACCTTTGCACAAATTT 98  
 Db 1 GGCANAGCGGAAGCACTGAAAGTCTCANATTCNNAATTTNAACTTTGCACAAATTT 60  
 Qy 99 CATACTAATTTTCTCAAGGAATTTGACAGCACTTCCCTGTGGGAAGGCCGCTCAAT 158  
 Db 61 AATCTAATTTTNNTCGAAGGAATTAAGCAAGCACTTCCCTGTGGGAAGGCCGCTCAAT 120  
 Qy 159 GCCCAATCAG-AGATTACCAAGAATCGCGATGTCTGCTGAAGAACATTTCA-GGGCA 216

Db 121 GCCCAATCAGAAATACCAAAATCGGCATGTATGCTGAAGAACATTCAGGGCA 180  
Qy 217 TGAAGCAGCTGGGCTCACCTGTGGAAGCACAGGAGGCTTGGAGCCCTGGAGAA-- 274  
Db 181 TGANGCAGCTGGGCTCACCTGTGGAAGCACAGGAGGCTTGGAGGCCCTGGAGG 240  
Qy 275 ---TGAATCAAGAGCTCCCGC-GGCTGGAGAACTTCTGCAGA-GACTTTGAGCTGC-- 327  
Db 241 ATTGNATTCAAGNCTCCCGCGGCTGGNGAACTTNTGCAGAGGACTTTGAGCTGNCA 300  
Qy 328 AGAAGGTGTGTACCTACCGCTCAACACCTT 358  
Db 301 GAAAGGTGTGTACCTACCGCTCAAAAANTT 331

RESULT 7  
US-09-783-590-3575  
; Sequence 3575, Application US/09783590  
; Patent No. US20020110850A1  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Patrick J.  
; APPLICANT: Haseltine, William A.  
; APPLICANT: Li, Haodong  
; APPLICANT: Rosen, Craig A.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
; FILE REFERENCE: PO-16.2C1  
; CURRENT APPLICATION NUMBER: US/09/783,590  
; CURRENT FILING DATE: 2000-02-15  
; PRIOR FILING DATE: 1995-04-12  
; PRIOR APPLICATION NUMBER: 08/420,856  
; PRIOR FILING DATE: 1994-11-21  
; NUMBER OF SEQ ID NOS: 12485  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3575  
; LENGTH: 356  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (32)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (42)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (45)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (51)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (61)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (73)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (75)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (134)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (145)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (168)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (216)

; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (246)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (267)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (276)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (280)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (320)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (325)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (335)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (338)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-3575

Query Match 22.0%; Score 168.8; DB 10; Length 356;  
Best Local Similarity 77.9%; Pred. No. 1.le-40;  
Matches 247; Conservative 0; Mismatches 57; Indels 13; Gaps 4;

Qy 39 GCCATGCGGGAAGCACTGAAAGTCTCATATTCGCCGAATTTTGAACCTTTGCACAAATTT 98  
Db 1 GGCAGAGCGGAAGCACTGAAAGTCTCATATNCCGAATTTNAACCTTTNCACAAATTT 60  
Qy 99 CATACTAATTTTCTCAAGAAATTTGACACAGACTTGCCCTGTGGGAAGCGCGCTCAAT 158  
Db 61 NATACTAATTTTNTNAAGAAATTAAGCAAGCACTTGCCCTGTGGGAAGCGCGCTCAAT 120  
Qy 159 GCCCAATCAGAGATTACCAAGATCGCGCATGT-CATGCTGAAGACATTCAGGGCAT 217  
Db 121 GCCCAATCAGAGNTTACCAAGANTCGGGATGTGCATGTGAAGANCATTCAGGGCAT 180  
Qy 218 GAAGCACCTGGCGCTCACCTGT-GGAAGCACAGGAGCGCTTGGAGGCCCTGGAGAAATG 276  
Db 181 GAGGCACCTGGGGTTCACCTGTGGGAAGCACAGCNAGGCGCTTGGAGGCCCTGGAGG 240  
Qy 277 GAATCAA-----GAGTCCCGGGGCTGGAGAACTTCTG-----CAGAGACTTTGAGCT 325  
Db 241 AATGNATTCAAGAGGTTCCCGGGGNTGGAGGAATTTTNTGCCAGAGAACTTTTGAAGCT 300  
Qy 326 GCAGAAAGGTGTGTACC 342  
Db 301 GCAGAAAGGTGTGTAAAC 317

RESULT 8  
US-09-783-590-3542  
; Sequence 3542, Application US/09783590  
; Patent No. US20020110850A1  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Patrick J.  
; APPLICANT: Haseltine, William A.  
; APPLICANT: Li, Haodong  
; APPLICANT: Rosen, Craig A.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
; FILE REFERENCE: PO-16.2C1  
; CURRENT APPLICATION NUMBER: US/09/783,590  
; CURRENT FILING DATE: 2000-02-15  
; PRIOR APPLICATION NUMBER: 08/420,856  
; PRIOR FILING DATE: 1995-04-12  
; PRIOR APPLICATION NUMBER: 08/346,731

;; PRIOR FILING DATE: 1994-11-21  
;; NUMBER OF SEQ ID NOS: 12485  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 3542  
;; LENGTH: 288  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (3)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (42)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (72)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (75)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (93)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (129)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (142)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (155)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (156)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (166)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (182)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (233)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (263)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (264)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (267)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (270)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (272)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (282)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; US-09-783-590-3542

Query Match 17.3%; Score 133.2; DB 10; Length 288;  
Best Local Similarity 85.8%; Pred. No. 3.8e-30;  
Matches 188; Conservative 0; Mismatches 27; Indels 4; Gaps 4;  
QY 46 CGGAAGCACTGAAAGTCTCAATTTCCGAACTTTTGAACCTTTTCACAAATTTTCATACTA 105  
|||||  
; Sequence 51, Application US/09764868  
; Patent No. US20020168711A1

Db 8 CGGAAGCACTGAAAGTCTCAATTTCCAAATTTNAAACCTTTTCACAAATTTTAAACTA 67  
QY 106 ATTTTCTCAAGGAAATTCAGCAAGCACTTGCCTGTGGGAAGCGCTCAAAATGCCCAAA 165  
|||||  
Db 68 ATTTNTTAAAGGAAATTAAGCAACGCTTGCCTGTGGAAAGCGCTCAAAATGCCCAAA 127  
QY 166 TCAGAGATTACCAAGAATCGCGCATGT-CATGCTGAAG-AACATTTCAGGCGCATGAAGCA 223  
|||||  
Db 128 TNA-AGATTACCAANAATCGCGCATGTNNATGCTGAAGNANCATTCAGGCGCATGNGCA 186  
QY 224 CTTGGCGGC-TCACCTCTGGAAGCACAGCGAGCGCTTGG 261  
|||||  
Db 187 CTTGGCGGCTTCACCTCTGGAAGCACAGCGAGCGCTTGG 225  
RESULT 9  
US-09-815-343-1490/c  
; Sequence 1490, Application US/09815343  
; Patent No. US2001005596A1  
; GENERAL INFORMATION:  
; APPLICANT: Meagher, Madeleine  
; APPLICANT: Xu, Jiangchun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
; FILE REFERENCE: 210121-504  
; CURRENT APPLICATION NUMBER: US/09/815,343  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1490  
; LENGTH: 426  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; US-09-815-343-1490

Query Match 15.7%; Score 120.8; DB 10; Length 426;  
Best Local Similarity 60.2%; Pred. No. 2.3e-26;  
Matches 219; Conservative 0; Mismatches 142; Indels 3; Gaps 1;  
QY 1 TCACCTCTGCTGTTTCAGAGCACAGTCAGCAAGAGGACGCCATGCCGGAAGCACTGAAAA 60  
|||||  
Db 365 TTACCGTGTGTTCCGACGCGCACTGTTGAAGGAGGACGCATGCCCTGCGACTCTGATGA 306  
QY 61 GTCTCATATTTCCGAAATTTTGAACCTTTGCACAAATTTTACATACTAATTTTCTCAAGAAA 120  
|||||  
Db 305 CGTGTCTTCTTCCAAACATCATCCCATCTATGATTTCCACAGAGGCTTCTCTGCGGAGG 246  
QY 121 TTGAGCAACGACTTGCCCTGTGGGAAGCGCGCTCAAAATGCCAAATCAGAG---ATTACC 177  
|||||  
Db 245 TGGAGCAGAGGCTGGCACTCTGGGAAGGCGCTTCCAAAGCCACACAAAAGGCGAGTCATC 186  
QY 178 AAAGAATCGGCGATGTCATGCTGAAGAACAATTCAGGCGCATGAAGCACCTGGCGGCTCACC 237  
|||||  
Db 185 AACGAATCGGGGACATCCTGCTCAGGAACATCGCCAGTAAAGGAGTTTACCAGCTACT 126  
QY 238 TGTGGAAGCACAGCGCGCTTGGAGGCCCTTGGAGATGAATCAAGAGCTCCCGGGCGG 297  
|||||  
Db 125 TCCAAAGACATGACGAGATCCTAACAGAACTGGAAAAGGCTACCAAACGCTGTGAAGAGT 66  
QY 298 TGGAGAACTTTCGACGAGACTTTGAGCTGCAGAAAGGTGTGTACTACCTACCGCTCAACACCT 357  
|||||  
Db 65 TGGAGGCACTGTACAGAGAGTTTGAGGCTGCGAAGAGTCTGCTACTTGCCTCTCAACACGT 6  
QY 358 TCCT 361  
|||||  
Db 5 TCCT 2

RESULT 10  
US-09-764-868-51  
; Sequence 51, Application US/09764868  
; Patent No. US20020168711A1

```
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 1718
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-51

Query Match      13.8%; Score 106; DB 9; Length 1718;
Best Local Similarity 64.9%; Pred. No. 1.3e-21;
Matches 157; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 528 CAGAAGCTGCACGAACTCAAGAAAGATTGATTGGCATTGACAATCTTGTGTTCCGGGA 587
Db 13 CAGAAGCTAAACGAGCTGCAGCGGACCTGTGGTCATAGAGAACCTCATTCTCTTGGC 72

QY 588 AGGGAGTTTCATCCGTCCTGGGACGCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGGCG 647
Db 73 AGGGAGTTTCATCCGTCGAGGGCTGCTTCACAAGCTCACCAGAAGGGCTCGCAGCAGAGG 132

QY 648 ATGTCTTCTCTCTTCAACGACGCTCTGCTATACACGAGCGGGGGCTGACGGCTCCCAAT 707
Db 133 ATGTCTTCTCTCTTCAACGACGCTCTGCTATACACGAGCGGGGGCTGACGGCTCCCAAT 192

QY 708 CAGTTAAAGTCCACGGGCGAGCTCCCGCTCTATGCATGACGATTGAGGAGCGGAAGAC 767
Db 193 CACTTCCGGATCCGGGCGCTCTTCCCTCCCAAGGCAATGCTGGTGAAGAAAGTGATAAC 252

QY 768 GA 769
Db 253 GA 254

RESULT 11
US-09-764-868-475
; Sequence 475, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 475
; LENGTH: 716
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (715)
; OTHER INFORMATION: n equals a,t,g, or c

Query Match      13.7%; Score 105.2; DB 9; Length 716;
Best Local Similarity 64.0%; Pred. No. 1.3e-21;
Matches 155; Conservative 2; Mismatches 85; Indels 0; Gaps 0;

QY 528 CAGAAGCTGCACGAACTCAAGAAAGATTGATTGGCATTGACAATCTTGTGTTCCGGGA 587
Db 3 CAGAAGCTWACGAGCTGCWKCGGACCTGTGGTCATAGAGAACCTCATTCTCTTGGC 62

QY 588 AGGGAGTTTCATCCGTCCTGGGACGCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGGCG 647
; (1510) (1510) (1510) (1510) (1510) (1510) (1510) (1510) (1510) (1510)
```

```
Db 63 AGGGAGTTTCATCCGTCGAGGGCTGCTTCACAAGCTCACCAGAAGGGCTTCGACGAGAGG 122
QY 648 ATGTCTTCTCTCTTCAACGACGCTCTGCTATACACGAGCGGGGGCTGACGGCTCCCAAT 707
Db 123 ATGTCTTCTCTCTTCAACGACGCTCTGCTATACACGAGCGGGGGCTTCGACGAGGACGAG 182
QY 708 CAGTTAAAGTCCACGGGCGAGCTCCCGCTCTATGCATGACGATTGAGGAGAGCGGAAGAC 767
Db 183 CACTTCCGGATCCGGGCGCTCTTCCCTCCCAAGGCAATGCTGGTGAAGAAAGTGATAAC 242

QY 768 GA 769
Db 243 GA 244

RESULT 12
US-09-764-868-48
; Sequence 48, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
; LENGTH: 2686
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (2505)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2569)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2644)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2645)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-48

Query Match      8.4%; Score 64.4; DB 9; Length 2686;
Best Local Similarity 47.8%; Pred. No. 3.9e-09;
Matches 287; Conservative 0; Mismatches 301; Indels 12; Gaps 3;

QY 175 ACCAAGAATCGGCGATGTCATGCTGAACACATTCAGGGCATGAAGCACCTTGGCGGCTC 234
Db 1505 ACCCAGCGCTCGGGACATCTCTGACAGAGCTGGCCCATCTCTGAAGATGTACGGCGAGT 1564

QY 235 ACCTGTGGAAGCACAGCGAGGCTTTGGAGCGCTTGAGAGATGGAATCAAGACTCCCGGC 294
Db 1565 ATGTCAGAACTTTGACCGAGCGTAGGCTGGTGAGCACGTGGACCCAGCGCTCCCGAC 1624

QY 295 GGCTGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTT---GTTACTTACCGCTCA 351
Db 1625 TGTTTAAAGACGTCRTCCACAGCATCCAGAGCAGGAGGTATGCGGAACCTTGACGCTGC 1684

QY 352 ACACCTTCTCTCTCGGCGCACTGCACCGGCTCATGCATACAGCAGGTCCTGGAGCGGC 411
Db 1685 AGCACCATGCTGGAGCGCGTGCAGAGGGTCCCGGTACGAGCTGTGCTCAAGACT 1744

QY 412 TGTGCAACACACCCCGCGAGCGACTTCAGGGACTGCGGAGCGCTTTGGGAG 471
Db 1745 ATCTGAAGAGGCTCCCGCAGGAGCGCCCGACCGGAAGGATCGCGAGAGGCTCTTTGGAGC 1804

QY 472 AGATCAGCGAGATGTTGGCAGAGCTCCACGGTACGATGATCAAGATGGAAGATTTCCAGA 531
; (1111) (1111) (1111) (1111) (1111) (1111) (1111) (1111) (1111) (1111)
```

Db 1805 TCATCTCCACAGCCGCCAACCACTCCAATGCTGCCAATTCGGAAAGTGAGAAAATGCACA 1864  
QY 532 AGCTGCAGCAACTCAAGAAAGATTTGATTGGCATTTGACAACTTCTTGTTCCGGGAAGGG 591  
Db 1865 ACCTTTGGAGTGTACAGACAGCTGGTGGGAAGAGACATTTGTCAACCCGGCCAATG 1924  
QY 592 AGTTCAATCGTCTGGCAGCCCTCAGCAAGCTCTCCGGGAAG-----GGGCTCCAGCAGC 645  
Db 1925 AACTGATCAAGGAGGCGCAAAATCCAGAAACTGTACGCCAAGAACGCCACCCCCAGGACC 1984  
QY 646 GCATGTTCTTCTTCAACAGAGCTCTGCTATACAGAGCGGGGGGCTGACGGCC---T 702  
Db 1985 GCACCTCTTCTGTTCAACAGCATGATCTTTACTGTGTGCCCAAGCTGCGGCTCATGG 2044  
QY 703 CCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGGCATGACGATGAGGAGAGCG 762  
Db 2045 GCCAAGTTACAGTCCGGGAGAGATGGACATCTCAGGCTCCAGGTGCAGGATATCG 2104

## RESULT 13

US-09-799-799-1  
; Sequence 1, Application US/09799799  
; Patent No. US20020132291A1  
; GENERAL INFORMATION:  
; APPLICANT: YE, Jane et al.  
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE  
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF  
; FILE REFERENCE: CL001157  
; CURRENT APPLICATION NUMBER: US/09/799,799  
; CURRENT FILING DATE: 2001-03-07  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 3380  
; TYPE: DNA  
; ORGANISM: Human  
US-09-799-799-1

Query Match 8.4%; Score 64.4; DB 10; Length 3380;

Best Local Similarity 47.8%; Pred. No. 4.e-09;  
Matches 287; Conservative 0; Mismatches 301; Indels 12; Gaps 3;

QY 175 ACCAAGAATCGCGCATGTCATGCTGAAGAACATTCAGGGCATGAAGCACTGGCGGCTC 234  
Db 1225 ACCCAGGCTCGGGACATCTCTGAGAGCTGGCCCCATCTCTGAGATGTACGGCGAGT 1284  
QY 235 ACCTGTGGAAGCACAGCGAGGCTTGGAGGCCCTGGAGATGGAATCAAGAGCTCCCGGC 294  
Db 1285 ATGTCAAGAACTTTGACCGAGCGGTAGGCTGGTGGACACGTGGACCCAGCGCTCCCGC 1344  
QY 295 GCTGGAGAACTTCTGACAGACTTTGAGCTGCAGAGGTGT--GTTACCTACCGTCA 351  
Db 1345 TGTTTAAAGACCTGCTCACAGCATCCAGACGAGGATGTCGGGAACCTTCAGCGTGC 1404  
QY 352 ACACCTTCTCTCGGGCCACTGCACCGCTCATGCACTACAAGCAGGTCTCTGGAGCGGC 411  
Db 1405 ASCACCATGCTGAGCGCGTGACAGGGTCCCGCGTACGAGTGTCTCAAGGACT 1464  
QY 412 TGTGAACACACACCCCGGAGCCACGCCGACTTCAGGAGCTGCCAGCGGCTTTGGCAG 471  
Db 1465 ATCTGAAGAGGCTCCCGAGGACGCCGCCAGCCGGAAGGTGCGGAGAGTCTTTGGAGC 1524  
QY 472 AGATCAGGAGATGTGGCAGAGCTCCAGGTACCATGATCAAGATGAGAGATTTCCAGA 531  
Db 1525 TCATCTCACAGCGGCCCAACCACTCCAATGCTGCGCATTCGGAAAGTGAGAAAATGCACA 1584  
QY 532 AGCTCAGCAACTCAAGAAAGATTTGATTGGCATTTGACAATCTTGTGTTCCGGGAAGGG 591  
Db 1585 ACCTTTGGAGTGTACGAGCAGCTGGTGGGAAGAGACATTTGTCAACCCGGCCAATG 1644  
QY 592 AGTTCAATCGTCTGGCAGCCCTCAGCAAGCTCTCCGGGAAG-----GGGCTCCAGCAGC 645  
Db 1645 ATCTGAAGAGGCTCCCGAGGACGCCGCCAGCCGGAAGGTGCGGAGAGTCTTTGGAGC 1524

Db 1645 AACTGATCAAGGAGGGCCAAATCCAGAAACTGTACGCCAAGAACGGACCCCGCCAGGACC 1704  
QY 646 GCATGTTCTTCTGTTCAACGAGCTCTGCTATACAGAGCGGGGGCTGACGGCC---T 702  
Db 1705 GCCACCTCTTCTGTTCAACGAGCATGATCTTTACTGTGTCGCCAAGCTGGCGTCAATGG 1764  
QY 703 CCAATCAGTTTAAAGTCCACGGGAGCTCCCGCTCTATGGCATGACGATTTGAGGAGAGCG 762  
Db 1765 GCCAAGTTACAGTCCGGGAGAGATGGACATCTCAGGCTCCAGGTGCAGGATATCG 1824

## RESULT 14

US-09-764-868-316  
; Sequence 316, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 316  
; LENGTH: 1091  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (1059)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-764-868-316

Query Match 8.3%; Score 64; DB 9; Length 1091;

Best Local Similarity 47.7%; Pred. No. 3.le-09;  
Matches 286; Conservative 1; Mismatches 301; Indels 12; Gaps 3;

QY 175 ACCAAGAATCGCGCATGTCATGCTGAAGAACATTCAGGGCATGAAGCACTGGCGGCTC 234  
Db 188 ACCCAGGCTCGGGACATCTCTGAGAGCTGGCCCCATCTCTGAGATGTACGGCGAGT 247  
QY 235 ACCTGTGGAAGCACAGCGAGGCTTGGAGGCCCTGGAGAATGGAATCAAGAGCTCCCGGC 294  
Db 248 ATGTCAAGAACTTTGACCGAGCGGTAGGCTGGTGGACACGTGGACCCAGCGCTCCCGC 307  
QY 295 GCTGGAGAACTTCTGACAGACTTTGAGCTGCAGAGAGTGT---GTTACCTACCGTCA 351  
Db 308 TGTTTAAAGAGCTCTCCACAGCATCCAGAACGAGGATGTCGGGAACCTTCAGCGTGC 367  
QY 352 ACACCTTCTCTCGGGCCACTGCACCGCTCATGCACTACAAGCAGGTCTCTGGAGCGGC 411  
Db 368 AGCACCATGCTGGAGCCGTGACAGAGGTCCCGGTACGAGCTGTCTCAAGGACT 427  
QY 412 TGTGAACACACACCCCGGAGCCACGCCGACTTCAGGAGCTGCCGAGCGGCTTTGGCAG 471  
Db 428 ATCTGAAGAGGCTCCCGAGGAGCGCCCGAGAGAGTTCGGAGAGGCTCTTTGGAGC 487  
QY 472 AGATCAGGAGATGTGGCAGAGCTCCAGGTACATGATCAAGATGAGAGATTTCCAGA 531  
Db 488 TCATCTCACAGCGGCCCAACCACTCCAATGCTGTCATTCGGAAAGTGAGAAAATGCACA 547  
QY 532 AGCTCAGCAACTCAAGAAAGATTTGATTGGCATTTGACAATCTTGTGTTCCGGGAAGGG 591  
Db 548 AGCTTTGGAGGTGTACGAGCAGCTGGGTGGGAAGAGACATTTGTCAACCCGGCCAATG 607  
QY 592 AGTTCAATCGTCTGGCAGCCTCAGCAAGCTCTCCGGGAAG-----GGGCTCCAGCAGC 645  
Db 608 AACTGATCAAGGAGGCCAAATCCAGAAACTGTACGCCAAGAACGCCACCCCGCCAGGACC 667  
QY 646 GCATGTTCTTCTGTTCAACGAGCTCTGCTATACAGAGCGGGGGCTGACGGCC---T 702  
Db 668 GCCACCTCTTCTGTTCAACAGCATGATCTTTACTGTGTGCCCAAGCTGCGGCTCATGG 727



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2002, 04:24:01 ; Search time 2433 Seconds  
(without alignments)  
5118.919 Million cell updates/sec

Title: US-09-555-342B-1\_COPY\_1733\_2501  
Perfect score: 769  
Sequence: 1 tcactctgtggttcagagc.....attgaggagagcgaagacga 769

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_Other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	744.8	96.9	1043	BQ072025	BQ072025 AGENCOURT
2	699.6	91.0	926	BQ706499	BQ706499 AGENCOURT
3	682	88.7	988	BQ674703	BQ674703 AGENCOURT
4	628.4	81.7	909	BQ646351	BQ646351 AGENCOURT
5	592	77.0	621	BG770181	BG770181 602744940
6	591	76.9	861	BG764061	BG764061 602737071

7	577.6	75.1	890	12	BE907778	BE907778
8	573.2	74.5	901	12	BG475554	BG475554
9	560.6	72.9	788	12	BG750463	BG750463
c 10	525.2	68.3	618	14	BQ331885	BQ331885
11	505.6	65.7	923	14	BQ950768	BQ950768
12	494.4	64.3	496	14	BQ705217	BQ705217
13	475.4	61.8	489	9	AL121548	AL121548
14	463	60.2	861	12	BG420356	BG420356
15	457	59.4	1026	13	BM551766	BM551766
16	444	57.7	454	14	BM728340	BM728340
17	414.8	53.9	896	14	BQ672632	BQ672632
18	412.4	53.6	757	10	BE282962	BE282962
19	398	51.8	719	12	BG248068	BG248068
20	366	47.6	449	10	BB840008	BB840008
21	340.4	44.3	741	12	BG122769	BG122769
22	337.4	43.9	351	14	D81819	D81819
c 23	336.8	43.8	358	12	BE812222	BE812222
24	326.8	42.5	466	14	RI4694	RI4694
25	323	42.0	516	13	BI898950	BI898950
26	322.6	42.0	475	14	BQ301106	BQ301106
c 27	314.6	40.9	876	14	BQ221203	BQ221203
28	304	39.5	414	10	BE271866	BE271866
29	304	39.5	516	13	BI535640	BI535640
30	303	39.4	1122	12	BE746268	BE746268
31	300.8	39.1	994	13	BI081073	BI081073
32	297.8	38.7	389	14	H11133	H11133
33	288.6	37.5	703	13	BJ059899	BJ059899
34	276	35.9	340	12	BF226117	BF226117
35	273.2	35.5	435	9	AI298217	AI298217
36	273.2	35.5	472	10	BE673853	BE673853
37	273.2	35.5	495	9	AI290936	AI290936
38	270	35.1	490	12	BF447678	BF447678
39	270	35.1	890	14	BQ681261	BQ681261
40	270	35.1	952	14	BQ676994	BQ676994
41	270	35.1	959	14	BQ679905	BQ679905
42	270	35.1	970	14	BQ676703	BQ676703
43	269.4	35.0	446	9	AI701115	AI701115
44	263.6	34.3	994	14	BQ683731	BQ683731
45	260	33.8	304	13	BI051634	BI051634

ALIGNMENTS

RESULT 1  
BQ072025  
LOCUS  
DEFINITION BQ072025 AGENCOURT\_6859787 NIH\_MGC\_47 Homo sapiens cDNA clone IMAGE:5928511  
5', mRNA sequence.  
ACCESSION BQ072025  
VERSION BQ072025.1 GI:19901071  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1043)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1CM2105 row: c column: 08  
High quality sequence stop: 626.  
Location/Qualifiers 1..1043

FEATURES  
source

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5928511"
/clone_lib="NIH_MGC_47"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 266 a 290 c 294 g 193 t
ORIGIN
Query Match 96.9%; Score 744.8; DB 14; Length 1043;
Best Local Similarity 99.7%; Pred. No. 4.7e-186;
Matches 746; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 9 TGGTTTCAGACAGTGCAGCAAGAGGAGCGCCATGCCGGAAGCACTGAAAAGTCTCAT 68
DB 1 TGGTTTCAGACAGTGCAGCAAGAGGAGCGCCATGCCGGAAGCACTGAAAAGTCTCAT 60
QY 69 TTCCCGAATTTTGAACCTTTGCACAAATTTTCATACATAATTTTCTCAAGGAAATTCAGCAA 128
DB 61 TTCCCGAATTTTGAACCTTTGCACAAATTTTCATACATAATTTTCTCAAGGAAATTCAGCAA 120
QY 129 CGACTTGGCCCTGTGGGAAGGCGGCTCAATGCCCAATCAGAGATTACCAAGAATCGGC 188
DB 121 CGACTTGGCCCTGTGGGAAGGCGGCTCAATGCCCAATCAGAGATTACCAAGAATCGGC 180
QY 189 GATGTCATGCTGAAAGCAATTCAGGGCATGAGACACTTGGGGCTCACCCTGTGGAGACAC 248
DB 181 GATGTCATGCTGAAAGCAATTCAGGGCATGAGACACTTGGGGCTCACCCTGTGGAGACAC 240
QY 249 ACGGAGGCTTGGAGGCGCTTGAGAAATGAATCAAGAGCTCCCGCGGCTGGAGAACTTC 308
DB 241 ACGGAGGCTTGGAGGCGCTTGAGAAATGAATCAAGAGCTCCCGCGGCTGGAGAACTTC 300
QY 309 TGCAGAGACTTTGAGCTGCAGAAAGTGTGTTACCTTACCGCTCAACACCTTCTCTCTGGG 368
DB 301 TGCAGAGACTTTGAGCTGCAGAAAGTGTGTTACCTTACCGCTCAACACCTTCTCTCTGGG 360
QY 369 CCAGTGCACCGGCTCATGCATACAGCAGTCTCTGGAGCGGCTGTGCAACACCAACCCG 428
DB 361 CCAGTGCACCGGCTCATGCATACAGCAGTCTCTGGAGCGGCTGTGCAACACCAACCCG 420
QY 429 CCGAGCCACGCGGCTTCCAGGAGCTGCGGAGCGGCTTTGGCAGAGATCAGCAGATGGTG 488
DB 421 CCGAGCCACGCGGCTTCCAGGAGCTGCGGAGCGGCTTTGGCAGAGATCAGCAGATGGTG 480
QY 489 GCACAGCTCCAGGCTACGATGATCAAGATGAGAAATTTCCAGAGCTGCACGAATCAAG 548
DB 481 GCACAGCTCCAGGCTACGATGATCAAGATGAGAAATTTCCAGAGCTGCACGAATCAAG 540
QY 549 AAAGATTTGATTTGATTCACAAATCTTGTGTTCCGGGAAGGAGTTTCATCGCTGGGC 608
DB 541 AAAGATTTGATTTGATTCACAAATCTTGTGTTCCGGGAAGGAGTTTCATCGCTGGGC 600
QY 609 AGCCTTCAGCAAGCTCTCGGGAAGGGCTCCAGCAGCGCATGTTCTCTCTGTTCAACAC 668
DB 601 AGCCTTCAGCAAGCTCTCGGGAAGGGCTCCAGCAGCGCATGTTCTCTCTGTTCAACAC 660
QY 669 GTCTGCTATACAGAGCGGGGGCTGACGGCCCTCCAACTAGTTTAAAGTCCACGGGAG 728
DB 661 GTCTGCTATACAGAGCGGGGGCTGACGGCCCTCCCACTAGTTTAAAGTCCACGGGAG 720
QY 729 CTCGGCTCTATGGCATGACGATTTAGG 756
DB 721 CTCGGCTCTATGGCATGACGATTTAGG 748
```

RESULT 2  
LOCUS B0706499 926 bp mRNA linear EST 16-JUL-2002  
DEFINITION AGENCOURT\_8474935 NIH\_MGC\_113 Homo sapiens cDNA clone IMAGE:6301634  
5', mRNA sequence.  
ACCESSION B0706499  
VERSION B0706499.1 GI:21845398  
KEYWORDS EST  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 926)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Dr. Mark Watson  
cDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1CM2517 row: n column: 03  
High quality sequence stop: 601.  
Location/Qualifiers  
1. 926  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6301634"  
/clone\_lib="NIH\_MGC\_113"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: spleen; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGGAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."  
BASE COUNT 242 a 247 c 254 g 183 t  
ORIGIN  
Query Match 91.0%; Score 699.6; DB 14; Length 926;  
Best Local Similarity 98.6%; Pred. No. 4e-174;  
Matches 716; Conservative 0; Mismatches 9; Indels 1; Gaps 1;  
QY 1 TCACCTTCGTGCTTTCAGAGCAGTGCAGCAAGAGGAGCGCCATGCCGGAAGCACTGAAAA 60  
DB 114 TCACCTTCGTGCTTTCAGAGCAGTGCAGCAAGAGGAGCGCCATGCCGGAAGCACTGAAAA 173  
QY 61 GTCTCATATTCGCCAATTTTGAACCTTTGCACAAATTTTCTACTACTAATTTTCTCAAGGAAA 120  
DB 174 GTCTCATATTCGCCAATTTTGAACCTTTGCACAAATTTTCTACTACTAATTTTCTCAAGGAAA 233  
QY 121 TTGAGCAACAGCACTTCCCTGTGGGAAGGCGCTCAATGCCCCAAATCAGAGATTACCAAAA 180  
DB 234 TTGAGCAACAGCACTTCCCTGTGGGAAGGCGCTCAATGCCCCAAATCAGAGATTACCAAAA 293  
QY 181 GAATTCGGCGATGTCATGCTGAAGAACATTCAGGCGCATGAAGCACTGGCGGCTCACCTGT 240  
DB 294 GAATTCGGCGATGTCATGCTGAAGAACATTCAGGCGCATGAAGCACTGGCGGCTCACCTGT 353  
QY 241 GGAAGCAGCAGCGGCGCTTGGAGGCGCTTGGAGATGGAATCAAGAGCTCCCGGGCGGTGG 300  
DB 354 GGAGCAGCAGCGGCGCTTGGAGGCGCTTGGAGATGGAATCAAGAGCTCCCGGGCGGTGG 413  
QY 301 AGAAGCTTTCAGAGACTTTCAGCTGCAAGAGGTGTGTTACCTACCGCTCAACACCTTCC 360  
|||||

Db 414 AGAATCTCTGCAGACATTTGAGCTGCAGAAAGGTGTGTTACCTACCGCTCAACACCTTCC 473  
QY 361 TCCTGCGGCACCTGCACCGGCTCATGCACCTACAAGCAGGTCTCTGGAGCGGTGTGCAAAAC 420  
Db 474 TCCTGGGCGCACTGCACCGGCTCATGCACCTACAAGCAGGTCTCTGGAGCGGTGTGCAAAAC 533  
QY 421 ACCACCGCGGAGCCAGCCGCACTTCAGGAGCTCCGAGCGCTTGGCAGAGATCAACGG 480  
Db 534 ACCACCGCGGAGCCAGCCGCACTTCAGGAGCTCCGAGCGCTTGGCAGAGATCAACGG 593  
QY 481 AGATGTCGACAGCTCCACCGTACGATCATCAAGATGAGAAATTCAGAGCTGCACG 540  
Db 594 AGATGTCGACAGCTCCACCGTACGATCATCAAGATGAGAAATTCAGAGCTGCACG 653  
QY 541 AACTCAAGAAAGATTTGATTTGGCATTTGACAAATCTTTGTGTTCCGGGAAGGGAGTTTCAATCC 600  
Db 654 AACTCAAGAAAGATTTGATTTGGCATTTGACAAATCTTTGTGTTCCGGGAAGGGAGTTTCAATCC 713  
QY 601 GTCTGGCAGCCTCAGCAAGCTCTCGGGAAGGGCTCCAGCAGCGCATGTTTCTTCCCTGT 660  
Db 714 GTCTGGCAGCCTCAGCAAGCTCTCGGGAAGGGCTCCAGCAGCGCATGTTTCTTCCCTGT 773  
QY 661 TCAACGAGCTCTCTATACAGAG - CCGGGGCTGACGGCTCCAATCAGTTTAAAGTC 719  
Db 774 TTCAGCAGCTCTCTATACAGAGCGCGGGGTGGACGCCCTCCAATCAGTTTAAAGTC 833  
QY 720 CACGGG 725  
Db 834 CACGGG 839

RESULT 3  
BO674703  
LOCUS  
DEFINITION BO674703 988 bp mRNA linear EST 15-JUL-2002  
AGENCOURT\_8188989 NIH\_MGC\_102 Homo sapiens cDNA clone IMAGE:6255217  
5', mRNA sequence.

ACCESSION BO674703  
VERSION BO674703.1 GI:21785537  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCM2405 row: p column: 02  
High quality sequence stop: 685.  
Location/Qualifiers  
1..988  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6255217"  
/clone\_lib="NIH\_MGC\_102"  
/tissue\_type="epidermoid carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: salivary gland; Vector: pOTB7; Site:1: XhoI;  
Site:2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAGCAG(G). Library constructed  
by Ling Hong in the laboratory of Gerald M. Rubin  
(University of California, Berkeley) using ZAP-cDNA  
synthesis kit (Stratagene) and Superscript II RT (Life

BASE COUNT 216 a 306 c 291 g 161 t 14 others  
ORIGIN  
Query Match 88.7%; Score 682; DB 14; Length 988;  
Best Local Similarity 99.6%; Pred. No. 1.8e-169;  
Matches 693; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
QY 75 AATTTTGAACCTTTTCACAAAATTTTCATATAATTTCTCAAGGAAATTTGAGCAACGACTT 134  
Db 1 AATTTTGAACCTTTTCACAAAATTTTCATATAATTTCTCAAGGAAATTTGAGCAACGACTT 60  
QY 135 GCCCTGTGGGAAGCGCTCAAAATCCCAAAATCAGAGATTTACCAAGAAATCGCGGATGTC 194  
Db 61 GCCCTGTGGGAAGCGCTCAAAATCCCAAAATCAGAGATTTACCAAGAAATCGCGGATGTC 120  
QY 195 ATGCTGAAGAACATTCAGGGCATGAGCACCTTGGCGGCTCACCTGTGGNAGCACAGCGAG 254  
Db 121 ATGCTGAAGAACATTCAGGGCATGAGCACCTTGGCGGCTCACCTGTGGNAGCACAGCGAG 180  
QY 255 GCCTTGGAGGCCCTTGAGAAATGGAATCAAGAGCTCCCGGGGCTGGAGAACTTCTGCGAGA 314  
Db 181 GCCTTGGAGGCCCTTGAGAAATGGAATCAAGAGCTCCCGGGGCTGGAGAACTTCTGCGAGA 240  
QY 315 GACTTTGAGCTGCAGAAAGGTGTGTTTACCTACCGCTCAACACCTTCTCTGCGGCCACTG 374  
Db 241 GACTTTGAGCTGCAGAAAGGTGTGTTTACCTACCGCTCAACACCTTCTCTGCGGCCACTG 300  
QY 375 CACCGGCTCATGCACCTACAGCAGGTCTCTGGAGCGGTGTGCAAAACACACCGCGCGAGC 434  
Db 301 CACCGGCTCATGCACCTACAGCAGGTCTCTGGAGCGGTGTGCAAAACACACCGCGCGAGC 360  
QY 435 CACCGGCTCATGCAGGACTTGCAGCGCTTGGCGAGAGATCACGGAGATGTGCGACAG 494  
Db 361 CACCGGCTCATGCAGGACTTGCAGCGCTTGGCGAGAGATCACGGAGATGTGCGACAG 420  
QY 495 CTCCACGGTACGATGATCAAGATGGAGAAATTTCCAGAGCTGCACGAACTCAAGAAAGAT 554  
Db 421 CTCCACGGTACGATGATCAAGATGGAGAAATTTCCAGAGCTGCACGAACTCAAGAAAGAT 480  
QY 555 TTGATTTGCATTTGACAAATCTTGTGTTCCGGGAAGGGAGTTTCCGTTGGGCGAGCTC 614  
Db 481 TTGATTTGCATTTGACAAATCTTGTGTTCCGGGAAGGGAGTTTCCGTTGGGCGAGCTC 540  
QY 615 AGCAAGCTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGACGCTCTG 674  
Db 541 AGCAAGCTCTCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCAACGACGCTCTG 600  
QY 675 CTATACACGAG - CCGGGGGCTGACGGCTTCCCAATCAGTTTAAAGTCCAGGGCAGCTCCC 733  
Db 601 CTATACACGAGNCCGGGGCTGACGGCTTCCCAATCAGTTTAAAGTCCAGGGCAGCTCCC 660  
QY 734 GCTCTATGGCATGACGATTTGAGGAGCGGAGACGA 769  
Db 661 GCTCTATGGCATGACGATTTGANGAGAGCGGAGACNA 696

RESULT 4  
BO646351  
LOCUS  
DEFINITION BO646351 909 bp mRNA linear EST 15-JUL-2002  
AGENCOURT\_8286354 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6299462  
5', mRNA sequence.  
ACCESSION BO646351  
VERSION BO646351.1 GI:21770523  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 909)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.  
Email: [cgabps-r@mail.nih.gov](mailto:cgabps-r@mail.nih.gov)  
Tissue Procurement: CGAP (Stanford)  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1CM2512 row: c column: 15  
High quality sequence stop: 589.

```

FEATURES
Source
    .909
Location/Qualifiers
1. .909
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="IMAGE:6299462"
   /clone_lib="NIH_MGC_100"
   /tissue_type="hepatocellular carcinoma, cell line"
   /lab_host="DH10B (phage-resistant)"
   /note="Organ: liver; Vector: pORF7; site_1: XhoI; Site-2:
ECORI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCACCGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
179 a      332 c      235 g      163 t
BASE COUNT
ORIGIN
```

Query Match	81.7%	Score 628.4	DB 14	Length 909
Best local Similarity	97.6%	Pred. No. 2.6e-155		
Matches 638	Conservative 0	Mismatches 16	Indels 0	Gaps 0
QY	116	GGAAATTTGAGCAACGACTTGGCCCTGTGGGAAGCGCGTCAATGCCAAATCAGAGATTA	175	
DB	1	GGAAATTTGAGCAACGACTTGGCCCTGTGGGAAGCGCGTCAATGCCAAATCAGAGATTA	60	
QY	176	CCAAAGAAATCGGCGATGTCTGTGAAGACATTTCAAGGCATGAAGCAACCTTGGCGGCTCA	235	
DB	61	CCAAAGAAATCGGCGATGTCTGTGAAGACATTTCAAGGCATGAAGCAACCTTGGCGGCTCA	120	
QY	236	CCTGTGGAAGCACAGCAGGCGCTTTGGAGGCCCTCGGAAATGGAATCAAGAGCTCCCGGCG	295	
DB	121	CCTGTGGAAGCACAGCAGGCGCTTTGGAGGCCCTCGGAAATGGAATCAAGAGCTCCCGGCG	180	
QY	296	GCTGGAGAACTTTCGACAGAGACTTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACAC	355	
DB	181	GCTGGAGAACTTTCGACAGAGACTTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACAC	240	
QY	356	CTTCCTCCTCGGGCCACTGCAACCGGCTCATGCACCTACAAGCAGTCTCTGGAGCGGCTGTG	415	
DB	241	CTTCCTCCTCGGGCCACTGCAACCGGCTCATGCACCTACAAGCAGTCTCTGGAGCGGCTGTG	300	
QY	416	CAACACACACCGCGGAGCCACGCGGACTTCAGGGACTCGGAGCGCGTTTGGCAGAGAT	475	
DB	301	CAACACACACCGCGGAGCCACGCGGACTTCAGGGACTCGGAGCGCGTTTGGCAGAGAT	360	
QY	476	CACGGAGATGGTGACACAGCTCCACGGTACGATGATCAAGATGGAGAAATTTCCAGAGAAGCT	535	
DB	361	CACGGAGATGGTGACACAGCTCCACGGTACGATGATCAAGATGGAGAAATTTCCAGAGAAGCT	420	
QY	536	GCAGAACTCAAGAAAGATTTGATTTGGATTTGCAATCTTGTGGTTTCGGGAAGGGAGTT	595	
DB	421	GCAGAACTCAAGAAAGATTTGATTTGGATTTGCAATCTTGTGGTTTCGGGAAGGGAGTT	480	
QY	596	CATCCGTCTGGCGAGCTTCAGCAAGCTCTCGGGGAAGGGCTCCACAGACGCGCATGTTCTT	655	
DB	481	CATCCGTCTGGCGAGCTTCAGCAAGCTCTCGGGGAAGGGCTCCACAGACGCGCATGTTCTT	540	
QY	656	CCTGTTTCAACGACGTCCTGCTATACAGAGCCGGGGGCTGACGGCCTCCAAATCAGTTTAA	715	

Db	541	CCTGTTCAACGAGCTCTGCTATACAGAGCGCGGGGCTGACGGCCTCCCAATCAGTTTAA	600
Qy	716	AGTCCACGGGCGAGCTCCCGCTCTATTGGCATGACGATTGAGGAGACGGAAGACGA	769
Db	601	AGCCCCGGCACCTCCCGCCTCTATGGCATGACCACTGACGACGCGAACCA	654
RESULT 5			
LOCUS	BG770181		
DEFINITION	602744940F1 NIH_MGC_49 Homo sapiens CDNA clone IMAGE:4878177 5',	621 bp mRNA linear	EST 15-MAY-2001
ACCESSION	BG770181		
VERSION	BG770181.1	GI:14080834	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 621)		
JOURNAL	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: <a href="mailto:cgabbs@email.nih.gov">cgabbs@email.nih.gov</a>		
	Tissue Procurement: ATCC/DCTD/DTF		
	cDNA Library Preparation: Ling Hong/Rubin Laboratory		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	<a href="http://image.llnl.gov">http://image.llnl.gov</a>		
	Plate: LLCMI763 row: 0 column: 10		
	High quality sequence stop: 621.		

[illegible]

QY 418 AACACACCCGCCGAGCCACCGCGACTTCAGGAGCTCCGAGCGCTTTGGCAGAGATCA 477  
|||||  
Db 242 AACACACCCGCCGAGCCACCGCGACTTCAGGAGCTCCGAGCGCTTTGGCAGAGATCA 301  
QY 478 CGGAGATGTTGGCAGACAGCTCCACGGTACATGATCAAGATGAGAGATTTCCAGAGCTGC 537  
Db 302 CGGAGATGTTGGCAGACAGCTCCACGGTACATGATCAAGATGAGAGATTTCCAGAGCTGC 361  
QY 538 AGCAACTCAAGAAAGATTTGATTGGCATTGACAATCTTGTGGTTCCGGGAAGGAGTTCA 597  
Db 362 AGCAACTCAAGAAAGATTTGATTGGCATTGACAATCTTGTGGTTCCGGGAAGGAGTTCA 421  
QY 598 TCCGCTGGGCGAGCCTCAGCAAGCTCTCCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCC 657  
Db 422 TCCGCTGGGCGAGCCTCAGCAAGCTCTCCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCC 481  
QY 658 TGTTCACAGCTCTGCTATACAGAGCGCGGGGCTGACGGCTCCCAATCAGTTTAAAG 717  
Db 482 TGTTCACAGCTCTGCTATACAGAGCGCGGGGCTGACGGCTCCCAATCAGTTTAAAG 541  
QY 718 TCCACGGGCGAGCTCCGCTCTATGCGATCAGCATGAGGAGCGGAAGACGA 769  
Db 542 TCCACGGGCGAGCTCCGCTCTATGCGATCAGCATGAGGAGCGGAAGACGA 593

RESULT 6  
BG764061  
LOCUS 602737071F1 NTH\_MGC\_49 Homo sapiens cDNA clone IMAGE:4862417 5',  
DEFINITION mRNA sequence.  
ACCESSION BG764061  
VERSION BG764061.1 GI:14074714  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 861)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC/DCTD/DTF  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LCM1722 row: n column: 18  
High quality sequence stop: 817.  
Location/Qualifiers  
1. .861  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4862417"  
/clone\_lib="NTH\_MGC\_49"  
/tissue\_type="melanotic melanoma, high MDR (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NTH\_MGC  
Library." 198 a 245 c 270 g 147 t 1 others

Query Match 76.9%; Score 591; DB 12; Length 861;  
Best Local Similarity 99.8%; Pred. No. 2e-145;  
Matches 591; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 178 AAAGAATCCGGCATGTCATGCTGAAGAACATTCAGGGCATGAAGCACTCGCGGGCTCACC 237  
|||||  
Db 2 AAAGAATCCGGCATGTCATGCTGAAGAACATTCAGGGCATGAAGCACTCGCGGGCTCACC 61  
QY 238 TGTGGGAAGCACAGCAGCGGCTTGGAGGCCCTTGAGAGATGAATCAAGAGCTCCCGCGGC 297  
|||||  
Db 62 TGTGGGAAGCACAGCAGCGGCTTGGAGGCCCTTGAGAGATGAATCAAGAGCTCCCGCGGC 121  
QY 298 TGSAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTGTTACCTACCGCTCAACACCT 357  
Db 122 TGSAGAACTTCTGCAGAGACTTTGAGCTGCAGAAAGTGTGTTACCTACCGCTCAACACCT 181  
QY 358 TCCTCTCTGGCGGCACTGCACCGGCTCATGCACCTACAAGCAGGTCTCTGGAGCGGCTGTGA 417  
|||||  
Db 182 TCCTCTCTGGCGGCACTGCACCGGCTCATGCACCTACAAGCAGGTCTCTGGAGCGGCTGTGA 241  
QY 418 AACACCACCGCGGAGCCACCGCGACTTCAGGGACTGCGAGCGCTTTGGCAGAGATCA 477  
|||||  
Db 242 AACACCACCGCGGAGCCACCGCGACTTCAGGGACTGCGAGCGGCTTTGGCAGAGATCA 301  
QY 478 CGGAGATGTTGGCAGACAGCTCCACGGTACGATGATCAAGATGGAAGATTTCCAGAACTGC 537  
Db 302 CGGAGATGTTGGCAGACAGCTCCACGGTACGATGATCAAGATGGAAGATTTCCAGAACTGC 361  
QY 538 AGCAACTCAAGAAAGATTTGATTGGCATTGACATTCACATCTTGTGGTTCCGGGAAGGAGTTCA 597  
|||||  
Db 362 AGCAACTCAAGAAAGATTTGATTGGCATTGACATTCACATCTTGTGGTTCCGGGAAGGAGTTCA 421  
QY 598 TCCGCTGGGCGAGCCTCAGCAAGCTCTCCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCC 657  
Db 422 TCCGCTGGGCGAGCCTCAGCAAGCTCTCCGGGAAGGGGCTCCAGCAGCGCATGTTCTTCC 481  
QY 658 TGTTCACAGCTCTGCTATACAGAGCGCGGGGCTGACGGCTCCCAATCAGTTTAAAG 717  
Db 482 TGTTCACAGCTCTGCTATACAGAGCGCGGGGCTGACGGCTCCCAATCAGTTTAAAG 541  
QY 718 TCCACGGGCGAGCTCCGCTCTATGCGATCAGCATGAGGAGCGGAAGACGA 769  
Db 542 TCCACGGGCGAGCTCCGCTCTATGCGATCAGCATGAGGAGCGGAAGACGA 593

RESULT 7  
BE907778  
LOCUS 601502061F1 NTH\_MGC\_70 Homo sapiens cDNA clone IMAGE:3903855 5',  
DEFINITION mRNA sequence.  
ACCESSION BE907778  
VERSION BE907778.1 GI:10401682  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 890)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLM9708 row: j column: 16  
High quality sequence stop: 646.  
Location/Qualifiers

FEATURES

```
source
1. .890
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3903855"
/clone_lib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/organ="pancreas; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.1 kb. Library constructed by Life Technologies."
BASE COUNT 251 a 242 c 241 g 156 t
ORIGIN
Query Match 75.1%; Score 577.6; DB 12; Length 890;
Best Local Similarity 98.4%; Pred. No. 7e-142; Indels 7; Gaps 7;
Matches 657; Conservative 0; Mismatches 4;
QY 104 TAATTTCTCAAGGAATAGCAACGACTTGCCTGTGGGAGCGCGCTCAAAATGCCCA 163
Db 1 TAATTTCTCAAGGAATAGCAACGACTTGCCTGTGGGAGCGCGCTCAAAATGCCCA 60
QY 164 ATATCAGAGATTACCAAGAAATCGGGATGCTATGCTGAAGAACATTCAGGGCATGAAGCA 223
Db 61 AATCAGAGATTACCAAGAAATCGGGATGCTATGCTGAAGAACATTCAGGGCATGAAGCA 120
QY 224 CTGG-CGGCTCACCTGTGGAAGCACAGCGAGCGCTTGGAGCGCCTGGAGATGGAATCA 282
Db 121 CTGGAGCGCTCACCTGTGGAAGCACAGCGAGCGCTTGGAGCGCCTGGAGATGGAATCA 180
QY 283 AGAGCTCCCGGGGC-TGGAGAACTTC-TGCAGAGACTTTGAGCTGCAGAAAGTGTGTTA 340
Db 181 AGAGCTCCCGGGCATGAGAACTTCATGTCAGAGACTTTGAGCTGCAGAAAGTGTGTTA 240
QY 341 CTTACCG-CTCAACAGCTTCCTCTCGGGCCACTGCACCGGCTCATGCACTACAAGCAGG 399
Db 241 CCTACCGCACTCAACAGCTTCCTCTCGGGCCACTGCACCGGCTCATGCACTACAAGCAGG 300
QY 400 TCTGGAGCGGCTGTGCAAAACACCGCCGAGCGACGTCGAGCTTCAGGGACTCCCGAG 459
Db 301 TCTGGAGCGGCTGTGCAAAACACCGCCGAGCGACGTCGAGCTTCAGGGACTCCCGAG 360
QY 460 CGGCTTTGCGAGATGACGGAGATGGTGGACAGCTCCACGGTACGATGATCAAGATGG 519
Db 361 CGGCTTTGCGAGATGACGGAGATGGTGGACAGCTCCACGGTACGATGATCAAGATGG 420
QY 520 AGAATTTCCAGAGCTGCACGAACTCAAGAAAGATTTGATTCGCATTCACATCTTGTGG 579
Db 421 AGAATTTCCAGAGCTGCACGAACTCAAGAAAGATTTGATTCGCATTCACATCTTGTGG 480
QY 580 TTCCGGGAAGGAGTTTCATCCGCTCTGGGAGCGCTCAGCAAGCTCTCGGGAAGGGCTCC 639
Db 481 TTCCGGGAAGGAGTTTCATCCGCTCTGGGAGCGCTCAGCAAGCTCTCGGGAAGGGCTCC 540
QY 640 AGCAGCGCATGTTCTCTCTGTTTCAA-CGACGTCTGCTATACAGCGCGGGG-CTGAC 697
Db 541 AGCAGCGCATGTTCTCTCTGTTTCAAACCGACGCTCTGCTATACAGCGCGGGGCTGAC 600
QY 698 GGCCTCCATCATGTTTAA-AGTCCAGCGGACGCTCCCGCTCTATGCATGACGATTGAGG 756
Db 601 GGCCTCCATCATGTTTAAAGTCCAGCGGACATCCCGCTCTATGCATGACGATTGAGG 660
QY 757 AGAGCGAA 764
Db 661 AGAGCGAA 668
RESULT 8
LOCUS BG475554
DEFINITION 602491544F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4620346 5', mRNA sequence.
ACCESSION BG475554
```

```
VERSION BG475554.1 GI:13407833
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 901)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1381 row: h column: 11
High quality sequence stop: 818.
FEATURES
source
Location/Qualifiers
1. .901
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4620346"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/organ="skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 232 a 243 c 253 g 173 t
ORIGIN
Query Match 74.5%; Score 573.2; DB 12; Length 901;
Best Local Similarity 89.5%; Pred. No. 1e-140;
Matches 664; Conservative 0; Mismatches 68; Indels 10; Gaps 4;
QY 1 TCACCTTCGTGGTTTCAGAGCACAGTGAGCAAGAGAGCGCCATGCGGAGCAGCTGAAAA 60
Db 155 TCACCTTCGTGGTTTCAGAGCACAGTGAGCAAGAGAGCGCCATGCGGAGCAGCTGAAAA 214
QY 61 GTCTCATATTCGCCGAATTTTGAACCTTTTGACAAAATTTTCACTAATTTTCTCAAGGAAA 120
Db 215 GTCTCATATTCGCCGAATTTTGAACCTTTTGACAAAATTTTCACTAATTTTCTCAAGGAAA 274
QY 121 TTGACACAGCACTTCGCTGTGGGAGGCGCTCAAAATGCCAAATCAGAGATTACCAAA 180
Db 275 TTGACACAGCACTTCGCTGTGGGAGGCGCTCAAAATGCCAAATCAGAGATTACCAAA 334
QY 181 GAATCGGCGATCTCATGCTGAAGAACATTCAGGGCATGAAGCACCTGGCGGCTCACCTGT 240
Db 335 GAATCGGCGATCTCATGCTGAAGAACATTCAGGGCATGAAGCACCTGGCGGCTCACCTGT 394
QY 241 GGAAGCACAGCGAGCGCTTGGAGGCGCTTGGAAGTGAATTAAGAGCTCCCGCGGCTGG 300
Db 395 GGAAGCACAGCGAGCGCTTGGAGGCGCTTGGAAGTGAATTAAGAGCTCCCGCGGCTGG 454
QY 301 AGAAGCTTTCGAGAGACATTTGAGCTGCAGAGAGTGTGTACCTACCGCTCACACCTTCC 360
Db 455 AGAAGCTTTCGAGAGACATTTGAGCTGCAGAGAGTGTGTACCTACCGCTCACACCTTCC 514
QY 361 TCCTCGGCGCACTGCACCGGCTCATGCACTACAAGCAGTCTCTGGAGCGGCTGTGCAAAAC 420
Db 515 TCCTCGGCGCACTGCACCGGCTCATGCACTACAAGCAGTCTCTGGAGCGGCTGTGCAAAAC 574
QY 421 ACCACCCCGGAGCACCGCCGACTTCA-GGGACTGCCGAGCGGCTTTGGCAGAGATCAG 479
```

```
Db 575 ACCAACCCGCGACGACCGCGACTTCAGGGGACTGGCGAGCGCTTTGCGAGAGATCACG 634
Qy 480 GAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATGAGAAATTTCCAGAGCTGCAC 539
Db 635 GAGATGGTGGCACAGCTCCACGGTACGATGATCAAGATGAGAAATTTCCAGAGCTGCAC 594
Qy 540 GAACCTCAAGAAAGATTTGATGGCATTTGATGATGATGATGATGATGATGATGATGATGAT 595
Db 695 AACTCAAGACAAAGATTTGATGGCATTTGATGATGATGATGATGATGATGATGATGATGAT 754
Qy 596 CATCCGCTGGCGACGCTCAGCAA----GCTCTGGGGAAGGGGCTCCAGCAGCGCATGT 651
Db 755 ATCCGGTCTGGCGACGCTTCAAGCAAGCTCCTCGGGGAAGGAGCTCCAGCAGCGCATGT 814
Qy 652 TCTTCTCTGTTCAACGAGCTCCT-GCTATACAGAGCGCGGGGCTGACGCGCTCCCAATCAG 710
Db 815 TCTTCTCTGTTCAACGAGCTCCTGCTGTTACAGAGCGCGGGGCTGACAGCGTCCCAATC 874
Qy 711 TTTAAAGTCCAGCGGAGCTCC 732
Db 875 CGTTAAAGCAAGGCGAGTCCC 896

RESULT 9
BG750463 788 bp mRNA linear EST 15-MAY-2001
LOCUS 602709323F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4845892 5',
DEFINITION mRNA sequence.
ACCESSION BG750463
VERSION BG750463.1 GI:14061116
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 788)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI685 row: n column: 05
High quality sequence stop: 756.
FEATURES
Location/Qualifiers
1..788
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4845892"
/tissue.lib="NIH_MGC_43"
/tissue.type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library. |"
BASE COUNT 196 a 210 c 229 g 153 t
ORIGIN

Query Match 72.9%; Score 560.6; DB 12; Length 788;
Best Local Similarity 94.7%; Pred. No. 2.le-137;
Matches 721; Conservative 0; Mismatches 24; Indels 16; Gaps 13;
```

```
Qy 24 GTGAGCAAGAGAGAGCCCATGCGGAGCAGCTCAAAA--GTCTCATATATCCCGAATTTGA 82
Db 1 GTGAGCAAGAGAGAGAGCCCATGCGGAGCAGCTGTCAACGCTCATATATCCCGAATTTGA 60
Qy 83 ACCTTTTGCACAAA--TTTCACTAATATTTTCTCAAGAAATTTGAGCAACGACTTGCCTGT 141
Db 61 ACCTTTTGCACAAAATATCATATATTTTCTCAAGAAATTTGAGCAACGACTTGCCTGT 120
Qy 142 GGAAGGCGCGCTCAA--ATGCCAAATCAGAGATTACCAAGAAATCGGC--GATGTCATGCT 199
Db 121 GGAAGGCGCGTCAAGATGCCAAATCAGAGATTACCAAGAAATCGGCTGATGTCATGCT 180
Qy 200 GAAGAACTTTCAGGGCATGAAGCACCTTGGCGTCACTTGTGGAAGCAGCAGGAGCGCTT 259
Db 181 GAAGAACTTTCAGGGCATGAAGCACCTTGGCGTCACTTGTGGAAGCAGCAGGAGCGCTT 240
Qy 260 GGAGGCGCTGGAGATGGAATCA--AGAGTCCGCGGGGCTGGAGAACTTCTTCGAGAGACT 318
Db 241 GGAGGCGCTGGAGATGGAATCAGAGAGCTCCGCGGGCTGGAGAACTTCTTCGAGAGACT 300
Qy 319 TTGAGCTGCAGAAAGTGTGTACCTA--CCGCTCAACACTTCTCCTCGGGCAGCTGCA 376
Db 301 TTGAGCTGCAGAAAGTGTGTACCTAGCCGCTCAACACTTCTCCTCGGGCAGCTGCA 360
Qy 377 CCGGCTCATGCACTA--CAAGCAGGCTCTGGAGCGGC---TGTGCAAAACACACCCGCCGA 432
Db 361 CCGGCTCATGCACTATCAAGCAGGCTCTGGAGCGGCTGTGGAAGACACACCGGCCCGCA 420
Qy 433 GCC-ACGCGCACTTCAGGAGCTGCCGAGCGCTTTGGCAGAGATCAGGAGATGGTGGA 491
Db 421 GCCGAGCGCGACTTTCAGGAGCTGCCGAGCGCTTTGGCAGAGATCAGGAGATGGTGGA 480
Qy 492 CAGCTCCACGCTACGATGATCAAGATGGAGAAATTTCCAGAACTGCAGAACTCAAGAAA 551
Db 481 CAGATCCACGCTACGATGATCAAGATGGAGAAATTTCCAGAACTGCAGAACTCAGAAA 540
Qy 552 GATTTGATTGGCATTTGACAATCTTGTGTTCCGG--GAAGGGAGTTTCCTCGTTCGGGAG 610
Db 541 GATTCGATTGGCATTTGACAATCTTGTGTTCCGGCGGAGGAGTTTCCTCGTTCGGGAG 600
Qy 611 CCTCAGCAAGCTCTCGGGGAA--GGGCTCCAGCAG--CGCATGTTCTTCTGTTCAACGAC 658
Db 601 ACTCAGCAAGCTCTCGGGGAA--GGGCTCCAGCAG--CGCATGTTCTTCTGTTCAACGAC 660
Qy 669 GTCTCTATACAGAGCGCGGGCTGACGGCTCCCAATCAGTTTAAAGTCCACGGGAG 728
Db 661 GTCTCTATACAGAGCGCGGGCTGACGG--CTCCAATCAGTTTAAAGTCCACGGGAG 719
Qy 729 CTCCCGCTCTATGGCATGACGATTGAGGAGCGGAAGACGA 769
Db 720 CTCCCGCTCTATGGCATGACGATTGAGGAGCGGAAGACGA 760

RESULT 10
BG331885/c 618 bp mRNA linear EST 17-MAY-2002
LOCUS BG331885/c
DEFINITION PM3-ET0268-100501-005-d11 ET0268 Homo sapiens cDNA, mRNA sequence.
ACCESSION BG331885
VERSION BG331885.1 GI:20973153
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 618)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
```

sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=PM3&t2=PM3-ET0268-  
100501-005-dl1&t3=2001-05-10&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 572.  
Location/Qualifiers  
1..618  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="ET0268"  
/dev\_stage="Adult"  
/note="Organ: lung\_tumor; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196.716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
BASE COUNT 128 a 162 c 159 g 169 t  
ORIGIN

Query Match 68.3%; Score 525.2; DB 14; Length 618;  
Best Local Similarity 99.1%; Pred. No. 4.2e-128;  
Matches 549; Conservative 0; Mismatches 3; Indels 2; Gaps 2;  
QY 3 ACTTGTGTTTTCAGACACAGTGTGACAAAGAGGAGCGCATCCGGAAGCACTGAAAGT 62  
DB 583 AATTCGTGTTTTCAGACACAGTGTGACAAAGAGGAGCGCATCCGGAAGCACTG-AAAGT 525  
QY 63 CTCATATCCCGAATTTGAACCTTTGACAAATTTCTATACATAATTTCTCAAGAAATT 122  
DB 524 CTCATATCCCGAATTTGAACCTTTGACAAATTTCTATACATAATTTCTCAAGAAATT 465  
QY 123 GAGCAACGACTTGCCTGTGGGAGGCGCTCAATGCCCAATCAGAGATTACCAAGA 182  
DB 464 GAGCAACGACTTGCCTGTGGGAGGCGCTCAATGCCCAATCAGAGATTACCAAGA 405  
QY 183 ATCGCGATGTCTGATGAAGAACATTTAGGCGATGAAGCACTGGCGGCTCACCTGTGG 242  
DB 404 ATCGCGATGTCTGATGAAGAACATTTAGGCGATGAAGCACTGGCGGCTCACCTGTGG 345  
QY 243 AAGCAGCGAGCGCTTGGAGCGCTTGGAGATGAAGATGAAGTCCGCGGCTGGAG 302  
DB 344 AAGCAGCGAGCGCTTGGAGCGCTTGGAGATGAAGATGAAGTCCGCGGCTGGAG 285  
QY 303 AACTTCTGAGAGACTTTGAGCTGAGAGGCTGTACCTACCGCTCAACACCTTCCTC 362  
DB 284 AACTTCTGAGAGACTTTGAGCTGAGAGGCTGTGTACCTACCGCTCAACACCTTCCTC 225  
QY 363 CTGCGGCCACTGCACCGGCTCATGCACTACAAGCAGGTCCTGGAGCGGCTGTGCAACAC 422  
DB 224 CTGCGGCCACTGCACCGGCTCATGCACTACAAGCAGGTCCTGGAGCGGCTGTGCAACAC 165  
QY 423 CACCGCGGAGCCACCGGACTTCAGGACTCCGAGCGGCTTTGGCAGAGATCAGGAG 482  
DB 164 CACCGCGGAGCCACCGGACTTCAGGACTCCGAGCGGCTTTGGCAGAGATCAGGAG 105  
QY 483 ATGGTGGCAGCTCCACGGTACGATGATCAAGATGGAGATTTCCAGAA-CCTGCACGA 541  
DB 104 ATGGTGGCAGCTCCACGGTACGATGATCAAGATGGAGATTTCCAGAAATGCTGCACGA 45

QY 542 ACTCAAGAAAGATT 555  
DB 44 ACTCAAGAAAGATT 31  
RESULT 11  
LOCUS BQ950768 923 bp mRNA linear EST 21-AUG-2002  
DEFINITION AGENCOURT\_8842056 Lupski\_sciatic\_nerve Homo sapiens cDNA clone  
IMAGE:6204242 5', mRNA sequence.  
ACCESSION BQ950768  
VERSION BQ950768.1 GI:22366246  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 923)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov  
Tissue Procurement: Dr. James R. Lupski  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM13625 row: d column: 03  
High quality sequence stop: 585.  
Location/Qualifiers  
1..923  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:6204242"  
/clone\_lib="Lupski\_sciatic\_nerve"  
/sex="male"  
/tissue\_type="sciatic nerve"  
/dev\_stage="adult, 70 yr"  
/lab\_host="DH10B"  
/note="Vector: pCMV-Sport6 (Life Technologies); Site\_1:  
NotI; Site\_2: SalI; cDNA made by oligo-dT priming.  
Directionally cloned using the following adaptors:  
5'-TCGACCATCGGCTCGG-3' and  
5'-GACTAGTCTAGTCGAGCGCGCCCT(15)-3'. Size selected >  
1 kb for average insert length 1.87 kb. This is a primary  
library, non-amplified. Library constructed by Life  
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
College of Medicine) and is available through Life  
Technologies."  
BASE COUNT 238 a 257 c 239 g 189 t  
ORIGIN  
Query Match 65.7%; Score 505.6; DB 14; Length 923;  
Best Local Similarity 95.0%; Pred. No. 7.8e-123;  
Matches 555; Conservative 0; Mismatches 24; Indels 5; Gaps 3;  
QY 1 TCACCTTCGTGGTTTCAGAGCACAGTGTGACAAAGAGAGCGCCATGCCGGAAGCACTGAAAA 60  
DB 316 TCACCTTCGTGGTTTCAGAGCACAGTGTGACAAAGAGAGCGCCATGCCGGAAGCACTGAAAA 375  
QY 61 GTCTCATATTCGGAAATTTGAACCTTTGCACAAATTTTCATACTAATTTTTCAGAGAAA 120  
DB 376 GTCTCATATTCGGAAATTTGAACCTTTGCACAAATTTTCATACTAATTTTTCAGAGAAA 435  
QY 121 TTGAGCAACGACTTGCCTGTGGAGGCGCTCAAAATGCCCAATTCAGAGATTACCAAA 180  
DB 436 TTGAGCAACGACTTGCCTGTGGAGGCGCTCAAAATGCCCAATTCAGAGATTACCAAA 495  
QY 181 GAATCGGCGATGTCATGCTGAAGAACATTCAGGCGCATGAAGCACCTTGGCGGCTCACCTGT 240  
DB 181 GAATCGGCGATGTCATGCTGAAGAACATTCAGGCGCATGAAGCACCTTGGCGGCTCACCTGT 240



```

Db 496 GAATCGGCGATGTCATGCTGAAGAACATTCAGGGCATGAAGACACCTGGCGGCTCACCTGT 555
QY 241 GGAAGCACAGCGAGCGCTTGGAGCCCTTGAGAAATGAATCAAGAGCTCCCGCGGCTGG 300
Db 556 GGAAGCACAGCGAGCGCTTGGAGCCCTTGAGAAATGAATCAAGAGCTCCCGCGGCTGG 615
QY 301 AGAACHTTCGAGAGACTTTGAGCTGCAGAGGTGTGTACCTACCGCTCAACACCTTCC 360
Db 616 AGAACHTTCGAGAGACTTTGAGCTGCAGAGGTGTGTACCTACCGCTCAACACCTTCC 675
QY 361 TCTCTGGCGCACTGCACCGGCTCATGCACTACAGCAGGTCTCTGGAGCGGCTGTGCAAC 420
Db 676 TCTCTGGCGCACTGCACCGGCTCATGCACTACAGCAGGTCTCTGGAGCGGCTGTGCAAC 735
QY 421 ACCACCGCGCG--AGCCAGCGGCTTACGGGACTCCGAGCGCG--TTTGGCAGAGATCA 477
Db 736 ACCACCGCGCGGAGCACGCCGACTTACGGGACTCCGAGCGGCTTTTGGCAGAGATCA 795
QY 478 C--GGAGATGGTGGCACAGCTCCACGGTAGGATGATCAGATGGAGAAATTTCCAGAAGCT 535
Db 796 CCGGAAATGGTGGCACAGCTCCACGGTAGGATGATCAGATGGAGAAATTTCCCGAAGCT 855
QY 536 GCACGAACTCAAGAAAGATTGATTGGCATTTGACAATCTTTGTGG 579
Db 856 GGCCCAACTCAGAAAGATTGATTGGCATTTGAACATCTCTGGG 899

RESULT 12
BM705217
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

BM705217 496 bp mRNA linear EST 28-FEB-2002
UI-E-Cil-aggf-c-04-0-UI.r1 UI-E-Cil Homo sapiens cDNA clone
BM705217
BM705217.1 GI:19018475
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 496)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse
Location/Qualifiers
1..496
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-Cil-aggf-c-04-0-UI"
/clone_lib="UI-E-Cil"
/tissue_type="RPE and Choroid"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/node="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-Cil is a normalized cDNA library containing the
following tissue(s): RPE and Choroid. The library was
constructed according to Bonaldo, Lennon and Soares,
```

```

Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(NT)18 tail. The sequence tag for this library is ACCTA.
This library was created for the program, Gene Discovery
in the Visual System, supported by National Eye Institute
(NEI)."
BASE COUNT 133 a 131 c 133 g 99 t
ORIGIN
Query Match 64.3%; Score 494.4; DB 14; Length 496;
Best Local Similarity 99.8%; Pred. No. 5.3e-120;
Matches 495; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 74 GAATTTTGAACCTTTGCACAAATTTTCATATAATTTTCTCAAGGAAATTTGAGCAAGCACT 133
Db 1 GAATTTTGAACCTTTGCACAAATTTTCATATAATTTTCTCAAGGAAATTTGAGCAAGCACT 60
QY 134 TGCCTCTGTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGAAATCCGCGATGT 193
Db 61 TGCCTCTGTGGGAAGCGCGCTCAATGCCCAATCAGAGATTACCAAGAAATCCGCGATGT 120
QY 194 CATGCTGAAGAACATTCAGGGCATGAAGCACCTGGCGGCTCACCTGTGGAAGCACAGCGA 253
Db 121 CATGCTGAAGAACATTCAGGGCATGAAGCACCTGGCGGCTCACCTGTGGAAGTACAGCGA 180
QY 254 GGCTTTGGAGGCCCTCGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAG 313
Db 181 GGCTTTGGAGGCCCTCGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTTCTGCAG 240
QY 314 AGACTTTGAGCTGCAGAAAGGTGTGTACCTACCGCTCAACACCTTCTCTCGGCGCACT 373
Db 241 AGACTTTGAGCTGCAGAAAGGTGTGTACCTACCGCTCAACACCTTCTCTCGGCGCACT 300
QY 374 GCACCGGCTCATGCACTACAGCAGGTCTGTGGAGCGGCTGTGCAACACACCCCGCGAG 433
Db 301 GCACCGGCTCATGCACTACAGCAGGTCTGTGGAGCGGCTGTGCAACACACCCCGCGAG 360
QY 434 CCACGCGGCTTTCAGGAGCTGCCGAGCGGCTTTGGCAGAGATCACGAGATGGTGGCACA 493
Db 361 CCACGCGGCTTTCAGGAGCTGCCGAGCGGCTTTGGCAGAGATCACGAGATGGTGGCACA 420
QY 494 GCTCCAGGTACGATGATCAAGATGGAGAAATTTCCAGAGCTGCAGAACTCAAGAAAGA 553
Db 421 GCTCCAGGTACGATGATCAAGATGGAGAAATTTCCAGAGCTGCAGAACTCAAGAAAGA 480
QY 554 TTTGATTGGCATTGAC 569
Db 481 TTTGATTGGCATTGAC 496

RESULT 13
AL121548
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AL121548 489 bp mRNA linear EST 25-FEB-2000
DKFP762LI158_r1 762 (synonym: hmel2) Homo sapiens cDNA clone
DKFP762LI158 5', mRNA sequence.
AL121548
AL121548.1 GI:5927549
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 489)
Ottewaelder, B., Oberwater, B., Mewes, H.W., Gassenhuber, J. and
Wiemann, S.
EST (Ottewaelder, et al.)
Unpublished (1999)
Contact: Ottewaelder B
```

MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Medigenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. No s1 sequence available.
This clone (DKFZp762L158) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de.
FEATURES
Location/Qualifiers
1..489
/organism="Homo sapiens"
/db\_xref="taxon:9606"
/clone="DKFZp762L158"
/clone\_lib="762 (synonym: hmel2)"
/tissue\_type="melanoma (MeWo cell line)"
/dev\_stage="adult"
/lab\_host="DH10B"
/note="Vector: pSport1; Site\_1: NotI; Site\_2: SalI"
BASE COUNT 121 a 134 c 143 g 91 t
ORIGIN
Query Match 61.8%; Score 475.4; DB 9; Length 489;
Best Local Similarity 99.6%; Pred. No. 5.5e-115;
Matches 487; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 125 GCACGACTTGCCCTGGGAGCGGCTCAATGCCCAATCAGAGATTACCAAGAAT 184
Db 1 GCAACGACTTGGCCTGGGAGCGGCTCAATGCCCAATCAGAGATTACCAAGAAT 60
QY 185 CGCGATGTCATGCTGAAGCAATTCAGGCGCATGAAGCACTGGCGCTCCTGTGGA 244
Db 61 CGCGATGTCATGCTGAAGCAATTCAGGCGCATGAAGCACTGGCGCTCCTGTGGA 120
QY 245 GCACGCGAGGCTTGGAGCGCTGGAGAAATGGAATCAAGAGCTCCCGCGCTGGAGAA 304
Db 121 GCACGCGAGGCTTGGAGCGCTGGAGAAATGGAATCAAGAGCTCCCGCGCTGGAGAA 180
QY 305 CTTCTGCAGAGCTTGACCTGCAGAGAGTGTGTACCTACCGCTCAACACCTTCTCTCT 364
Db 181 CTTCTGCAGAGCTTGTAGCTGCAGAGAGTGTGTACCTACCGCTCAACACCTTCTCTCT 240
QY 365 GGGCCACTGCACCGGCTCATGCACATCAAGCAGGTCCTGGAGCGGCTGTGCAACACCA 424
Db 241 GCGGCCACTGCACCGGCTCATGCATCAAGCAGGTCCTGGAGCGGCTGTGCAACACCA 300
QY 425 CCGCGCGACCGCGGACTTCAGGAGCTTCCAGCGGCTTTGGCAGAGATCACGAGAT 484
Db 301 CCGCGCGACCGCGGACTTCAGGAGCTTCCAGCGGCTTTGGCAGAGATCACGAGAT 360
QY 485 GGTGGCACAGCTCCAGGTCACGATGATCAAGATGAGAAATTTCCAGAGCTGCAGAACT 544
Db 361 GGTGGCACAGCTCCAGGTCACGATGATCAAGATGAGAAATTTCCAGAGCTGCAGAACT 420
QY 545 CAAGAAGATTTGATTGGCATTGACAACTCTGTGTTCCGGGAAGGAGTTCA-TCCGTC 603
Db 421 CAAGAAGATTTGATTGGCATTGACAACTCTGTGTTCCGGGAAGGAGTTCA-TCCGTC 480
QY 604 TGGCGAGCC 612
Db 481 TGGCGAGCC 489
RESULT 14
BG420356
LOCUS
DEFINITION 602448377F1 NIH\_MGC\_14 Homo sapiens cDNA clone IMAGE:4586971 5', mRNA linear EST 14-MAR-2001
ACCESSION BG420356
VERSION BG420356.1 GI:13326862
KEYWORDS EST.

human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 861)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DPF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLCM1316 row: i column: 20
High quality sequence stop: 569.
FEATURES
Location/Qualifiers
1..861
/organism="Homo sapiens"
/db\_xref="taxon:9606"
/clone="IMAGE:4586971"
/clone\_lib="NIH\_MGC\_14"
/tissue\_type="renal cell adenocarcinoma"
/lab\_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 188 a 249 c 268 g 156 t
ORIGIN
Query Match 60.2%; Score 463; DB 12; Length 861;
Best Local Similarity 100.0%; Pred. No. 1.4e-111;
Matches 463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 307 TCTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACACCTTCTCTCTGC 366
Db 3 TCTGCAGAGACTTTGAGCTGCAGAAAGTGTGTACCTACCGCTCAACACCTTCTCTCTGC 62
QY 367 GGCCTACTGACCGGCTCATGCACTCAACAGAGTCTTGGAGCGGCTGTGCAACACCA 426
Db 63 GGCCTACTGACCGGCTCATGCACTCAACAGAGTCTTGGAGCGGCTGTGCAACACCA 122
QY 427 CGCCGAGCCACCGGCTTTCAGGAGCTGCGGAGCTGCGGAGCTTGGCAGAGATCACGGAGATGG 486
Db 123 CGCCGAGCCACCGGCTTTCAGGAGCTGCGGAGCTGCGGAGCTTGGCAGAGATCACGGAGATGG 182
QY 487 TGGCAGAGCTCCACCGTACGATGATCAAGATGGAATTTCCAGAGCTGCAGAACTCA 546
Db 183 TGGCAGAGCTCCACCGTACGATGATCAAGATGGAATTTCCAGAGCTGCAGAACTCA 242
QY 547 AGAAGATTTGATTGGCATTGACAACTCTTGTGTTCCGGGAAGGAGTTTCATCCGTCG 606
Db 243 AGAAGATTTGATTGGCATTGACAACTCTTGTGTTCCGGGAAGGAGTTTCATCCGTCG 302
QY 607 GCAGCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCGGCTTCCAGCGCATGTTCTTCCGTTCAAGC 666
Db 303 GCAGCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCGGCTTCCAGCGCATGTTCTTCCGTTCAAGC 362
QY 667 AGCTCCTGCTATACACGAGCGGGGCTCACGCGCTTCCCAATCAGTTTAAAGTCCACGGGC 726
Db 363 AGCTCCTGCTATACACGAGCGGGGCTCACGCGCTTCCCAATCAGTTTAAAGTCCACGGGC 422
QY 727 AGCTCCCGCTCTATGGCATGACGATTGAGGAGGAGCAAGCA 769
Db 423 AGCTCCCGCTCTATGGCATGACGATTGAGGAGGAGCAAGCA 465

## RESULT 15

BM551766  
LOCUS BM551766 1026 bp mRNA linear EST 20-FEB-2002  
DEFINITION AGENCOURT\_6575294 NIH\_MGC\_98 Homo sapiens cDNA clone IMAGE:5479350  
5', mRNA sequence.  
ACCESSION BM551766  
VERSION BM551766.1 GI:18789071  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 1026)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LCM2000 row: h column: 07

High quality sequence stop: 675.

## FEATURES

source

Location/Qualifiers

1..1026

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5479350"

/clone\_lib="NIH\_MGC\_98"

/tissue\_type="astrocytoma grade IV, cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: brain; Vector: pOTB7; Site\_1: XhoI; Site\_2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GCGACGAG(G). Library constructed by Ling Hong

in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH\_MGC Library."

227 a 309 c 296 g 194 t

BASE COUNT

ORIGIN

Query Match 59.4%; Score 457; DB 13; Length 1026;  
Best Local Similarity 100.0%; Pred. No. 5.8e-110;  
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 313 GAGACTTTGAGCTGAGAGGTGTGTACCTACCGCTCAACACCTTCCTCTGGCGGCAC 372

Db |||||

1 GAGACTTTGAGCTGAGAGGTGTGTACCTACCGCTCAACACCTTCCTCTGGCGGCAC 60

QY 373 TGCACCGGCTCATGCACATACAGCAGGTCTCTGGAGCGGTGTGCAAAACACCAACCCCGCA 432

Db |||||

61 TGCACCGGCTCATGCACATACAGCAGGTCTCTGGAGCGGTGTGCAAAACACCAACCCCGCA 120

QY 433 GCCACCGGCTTCAGGAGCTCCGAGCGCTTTGGCAGAGATCACGGAGATGTTGGCAGC 492

Db |||||

121 GCCACCGGCTTCAGGAGCTCCGAGCGCTTTGGCAGAGATCACGGAGATGTTGGCAGC 180

QY 493 AGCTCCAGGTCAGATGATCAAGATGGAGATTTCCAGAGCTGCACGAACCTCAAGAAG 552

Db |||||

181 AGCTCCAGGTCAGATGATCAAGATGGAGATTTCCAGAGCTGCACGAACCTCAAGAAG 240

QY 553 ATTTGATTGGCAATTCACAATCTTGTGGTCCGGGAAGGAGTTTCATCCGCTCGGCAGCC 612

Db |||||

241 ATTTGATTGGCAATTCACAATCTTGTGGTCCGGGAAGGAGTTTCATCCGCTCGGCAGCC 300

QY 613 TCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCACACGCTCC 672

Db 301 TCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTCTGTTCACACGCTCC 360  
QY 673 TGCTATACACGAGCGGGGGCTGACGGCTCCAATCAGTTTAAAGTCCACGGGCGAGCTCC 732  
Db 361 TGCTATACACGAGCGGGGGCTGACGGCTCCAATCAGTTTAAAGTCCACGGGCGAGCTCC 420  
QY 733 CGCTCTATGGCATGACGATTTGAGGAGAGCGAAGACGA 769  
Db 421 CGCTCTATGGCATGACGATTTGAGGAGAGCGAAGACGA 457

Search completed: December 12, 2002, 06:39:28

Job time : 2443 secs

**THIS PAGE BLANK (USPTO)**